Computers in Biology and Medicine 178 (2024) 108703

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Diffuse tumors: Molecular determinants shared by different cancer types

Xuan Li [a,b], Dingyun Liu [a], Zhipeng Wu [a], Ying Xu [b,*]

[a] Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, College of Computer Science and Technology, Jilin University, Changchun, 130012, China
[b] School of Medicine, Southern University of Science and Technology, Shenzhen, China

ARTICLE INFO

Keywords:
Cancer morphology
Sialic acid
Circulating growth factors
Cancer occurrence risk
Cancer malignancy

ABSTRACT

Most cancer types have both diffuse and non-diffuse subtypes, which have rather distinct morphologies, namely scattered tiny tumors vs. one solid tumor, and different levels of aggressiveness. However, the causes for forming such distinct subtypes remain largely unknown. Using the diffuse and non-diffuse gastric cancers (GCs) as the illustrative example, we present a computational study based on the transcriptomic data from the TCGA and GEO databases, to address the following questions: (i) What are the key molecular determinants that give rise to the distinct morphologies between diffuse and non-diffuse cancers? (ii) What are the main reasons for diffuse cancers to be generally more aggressive than non-diffuse ones of the same cancer type? (iii) What are the reasons for their distinct immunoactivities? And (iv) why do diffuse cancers on average tend to take place in younger patients? The study is conducted using the framework we have previously developed for elucidation of general drivers cancer formation and development. Our main discoveries are: (a) the level of (poly-) sialic acids deployed on the surface of cancer cells is a significant factor contributing to questions (i) and (ii); (b) poly-sialic acids synthesized by *ST8SIA4* are the key to question (iii); and (c) the circulating growth factors specifically needed by the diffuse subtype dictate the answer to question (iv). All these predictions are substantiated by published experimental studies. Our further analyses on breast, prostate, lung, liver, and thyroid cancers reveal that these discoveries generally apply to the diffuse subtypes of these cancer types, hence indicating the generality of our discoveries.

1. Introduction

It has been well documented that some cancer types, such as stomach, breast (BCs), liver (HCs), lung (LCs), prostate (PCs), and thyroid (TCs) cancer, each consist of two morphologically distinct subtypes: diffuse and non-diffuse subtypes with the former each exhibiting multiple dispersed small tumors with poorly defined or absent glandular structures, whereas the latter each presents a single cohesive tumor [1–7]. In addition, the diffuse tumors generally have worse prognosis compared to their matching non-diffuse tumors [2,8–14]. While diffuse tumors have been reported at least as early as 1960s [15], very little is known about their underlying causes, particularly common causes shared by diffuse tumors across different cancer types.

Our comprehensive search of the public databases led to the identification of six cancer types, each having transcriptomic data for at least three tissue samples for both diffuse and non-diffuse tumors, respectively. Table 1 summarizes the detailed information of these cancer types: BCs [1,2], HCs [4], LCs [2,5], PCs [2,6], GCs [3,16], and TCs [7].

The goal of this study is to elucidate the distinct molecular factors that can explain all the distinct phenotypes associated with the two subtypes for each cancer type, through analyzing and mining the available transcriptomic data of the relevant tissues. Computational methods have been widely used to classify cancers [17,18], even cancer samples with limited annotations [19], including deep CNN coupled with the particle swarm optimization algorithm [20], extreme learning machines [21], and integrated machine learning frameworks [22–24]. However, our preliminary analyses have revealed that machine learning or AI techniques alone are yet to be able to identify molecular features that can well explain the distinct characteristics between the two subtypes of each relevant cancer type, as that requires logical reasoning using chemistry- and physics-based knowledge that bridges gaps among the information.

Our review of the literature reveals that signet-ring cell carcinoma (SRCC) is the predominant diffuse or diffuse-like subtype present in various cancer types, characterized by the presence of more than 50 % signet-ring cells within the tumor [2]. In addition, they generally have worse prognoses compared to the other tumors of the same cancer types. Among all such tumors, DGC is the most studied, largely due to two

* Corresponding author.
E-mail address: xuy9@sustech.edu.cn (Y. Xu).

https://doi.org/10.1016/j.compbiomed.2024.108703
Received 5 December 2023; Received in revised form 2 May 2024; Accepted 1 June 2024
Available online 2 June 2024

Table 1
Diffuse-like tumors of the six cancer types.

Cancer types	Information of diffuse-like tumors
GC	**Diffuse gastric cancer (DGC)** is characterized by the presence of multiple small, separate tumors that infiltrate the tissue stroma. DGC cells are poorly or un- differentiated and lack cell-to-cell adhesion. Glandular structures are rarely seen in DGC. Some of DGC tissues contain more than 50% signet-ring cells, referred to as signet-ring cell carcinoma. DGCs account for 32% of all GCs [3], and their average 5-year survival rate is lower than the other GCs, 45.6% vs. 57.7% [8].
BC	There are two types of diffuse-like BCs, accounting for < 3% of all BCs with significantly lower 5-year survival rates than other BCs, 40.5% vs. 63.2% for IBC and 65.1% vs. 90.0% for SRCC [2,9]. **Inflammatory breast cancer (IBC)** is the most aggressive BC subtype. Compared to other BCs, IBC is associated with acute inflammation and generally poorly differentiated with diffuse brawny infiltration, showing no underlying mass structure [1,9,25]. **Signet-ring cell carcinoma (SRCC)** is defined, according to the WHO's classification, as a poorly cohesive carcinoma in which more than 50% of tumor cells are with prominent cytoplasmic mucin and a crescent-shaped nucleus eccentrically placed. They are poorly or un-differentiated and each tumor appears as individual cells or in loose cell-clusters [2,26].
PC	There are two types of diffuse-like PCs, accounting for < 6% of all PCs with significantly lower 5-year survival rates than other PCs, < 65% vs. > 90% for those with high GG and 83.6% vs. 93.4% for SRCC [2,10]. PCs with **Gleason grade** (GG, from 1 to 5) ≥ 4 are defined as poorly or un-differentiated tumors growing in diffuse patterns. Tumors with GG 4 have poorly formed glandular lumina, while those in GG 5 are without glandular differentiation. Therefore, a tumor with Gleason score (GS, sum of the primary and secondary GG, range from 2 to 10) ≥ 8 with GGs ≥ 4 is considered as a diffuse-like tumor [6,10]. SRCC cancer tissues are derived from prostate epithelial cells [2,26].
LC	There are two types of diffuse-like LCs, accounting for < 15% of all LCs with significantly lower 5-year survival rates than other LCs, 7% vs. 18% for SCLC and 9.7% vs. 26.1% for SRCC [2,11,12]. **Small-cell lung carcinoma (SCLC)** is defined as tumors with cells having a relatively small size, a round-to-fusiform shape, scant cytoplasm, and absent or inconspicuous nucleoli. The tumors are mostly poorly or un- differentiated, fall short of glandular differentiation, and grow as diffuse sheets [5,27]. **SRCC** cancer tissues are derived from lung epithelial cells [2,26].
HC	**Diffuse hepatocellular carcinoma (DHCC)** also known as cirrhotomimetic HCC or cirrhosis-like HCC, is a subtype of HC. Different from massive HCC, DHCC shows cirrhosis-like diffuse growth. Cancer tissues exhibit pseudo-glandular and trabecular patterns and may lack well-demarcated boundaries [4,28,29]. DHCC generally blend into the background of the cirrhotic liver. It accounts for < 20% of all HCC [28,29] with significantly lower 5-year survival rates than other HCCs, 13.6% vs. 46.0% [13].
TC	**Diffuse sclerosing variant papillary thyroid carcinoma (DSVPTC)** is characterized by histologic features of numerous psammoma bodies, extensive lymphocytic infiltration, squamous metaplasia, diffuse fibrosis, calcification, and absence of string colloids together. DSVPTC, with highly diffuse appearance, can involve the thyroid gland extensively without forming a dominant mass [7,30]. This subtype accounts for < 7% of all TCs [7] with significantly lower 5-year survival rates than the classic papillary thyroid carcinoma, 74.4% vs. 89.4% [14].

reasons: (i) DGC tumors account for a large fraction, at least 32% of all GCs, and (ii) tumor-resection surgery is effective for treating this tumor subtype, making its tissue samples more available for biological research, compared to other diffuse-like subtypes where surgery is less applicable. Our study initially focuses on the two major subtypes of GCs, namely DGC and the intestinal gastric cancer (IGC), a typical non-diffuse subtype, accounts for 54% of all GC cases [3]; and then extends to the five other cancer types.

DGCs each consist of numerous disjoint tiny tumors infiltrating the tissue stroma while IGCs forms one solid tumor with strong cell-cell adhesion [16]. Published data indicate that DGCs tend to take place in younger patients and more female patients than IGCs [31]. Their progression tends to be more aggressive and easier to metastasize, having considerably lower survival rates than IGCs [16,32]. Some differences have been detected at the genomic and transcriptomic levels between DGCs and IGCs, including that (a) DGC tends to have reduced cell-cell adhesion along with mutations in adhesion genes like *CDH1* [33] and *RHOA* [33,34]; and (b) DGCs generally have lower levels of genomic instability compared to IGCs [35]. However, it remains largely unknown of how changes in these genes contribute to the distinct phenotypes and clinical behaviors of DGCs vs. IGCs.

We have previously developed a framework for elucidation of general and fundamental drivers of cancer onset and development. Briefly, the onset of a cancer is the result of chronic inflammation in the host organ, coupled with iron overload at the disease site. Once their levels reach beyond certain thresholds, Fenton reaction: $Fe^{2+} + H_2O_2 \rightarrow Fe^{3+} + \cdot OH + OH^-$ will take place without involving any enzyme, which has been widely observed in cancer tissues [36]. We have demonstrated statistically with strong supports from published experiments that cancer tissues in the TCGA database, including GC, all harbor Fenton reactions in cytosol, using superoxide ($O_2^{\bullet -}$) as the reducing molecule to convert Fe^{3+} back to Fe^{2+}, hence giving rise to persistent Fenton reactions in the following form, also referred to as the Haber-Weiss reaction:

$$O_2^{\cdot -} + H_2O_2 \xrightarrow{Fe^{2+}} O_2 + \cdot OH + OH^- \quad (1)$$

with Fe^{2+} as the catalyst [36]. We have further established via computational modeling that (a) the persistent production of OH^- by Fenton reactions can quickly saturate the intracellular pH buffer, hence driving the pH up and killing the host cells if not neutralized; (b) proton pumps or equivalents could not resolve the issues in a sustained manner since that will violate either the electroneutrality of the host cells or homeostasis of other ions [37]; and (c) multiple reprogrammed metabolisms (RMs), such as *de novo* biosynthesis of nucleotides and over-production and deployment of sialic acids (SAs) are induced each to produce numerous H^+ to neutralize the Fenton reaction-produced OH^- and keep the intracellular pH stable [36–38]. SAs are nine-carbon sugars, used as the capping molecule of glycan. They have long been known to be associated with poor prognosis of a cancer [39]. Our previous study has provided strong evidence that they play key roles in driving cancer migration and metastasis [40].

The main contribution of our study is to have identified key molecular determinants of the key characteristics between the diffuse and non-diffuse subtypes in different cancer types and to have explained statistically the key characteristics of the diffuse vs. non-diffuse subtypes, namely differences in morphology, clinical behaviors, and prognosis, in terms of the identified molecular factors, derived through computational analyses and modeling of the relevant transcriptome data.

Technically, we have combined machine learning technology with our previous established framework for cancer driver studies [36,38] to derive a mechanic explanation that links phenotypic differences to differences in chemical and physical microenvironments between diffuse and non-diffuse subtypes of cancers. We anticipate that this approach offers a novel and general perspective on distinguishing between diffuse and non-diffuse cancers, and yields valuable insights for enhancing cancer diagnosis, treatment, and prognosis.

2. Materials and methods

2.1. Data

2.1.1. Gene expression data

258 samples of RNA-seq data (Read-count in TPM values) of GC are downloaded from the TCGA database (https://portal.gdc.cancer.gov/) [41], of which 63 are DGC, 163 are IGC, and 32 are normal tissues (denoted as SNTs). RNA-seq data of diffuse-like and non-diffuse samples of BC (17 diffuse-like and 100 non-diffuse samples), PC (182 diffuse-like and 44 non-diffuse samples), HC (3 diffuse-like and 8 non-diffuse samples), and TC (4 diffuse-like and 36 non-diffuse samples), plus their corresponding controls, are also downloaded from TCGA. Microarray data of diffuse-like and non-diffuse LC samples (21 diffuse-like and 16 non-diffuse samples) are downloaded from the GEO (https://www.ncbi.nlm.nih.gov/geo/) dataset (GSE40275) [42]. RNA-seq data of seven red-blood cell samples are downloaded from GEO (GSE108378 and GSE63703) and used in our comparative analyses since red blood cells are known to employ the highest amount of SAs among all human cell types [40,43,44]. Details of these data are summarized in Supplementary Table S1.

To enable comparisons across different datasets, we have used the expression data of 18,504 protein-coding genes commonly shared by all the samples mentioned above.

2.1.2. Survival data

Kaplan-Meier survival curves of integrin genes and SA-related genes are downloaded from the UALCAN server (http://ualcan.path.uab.edu/) [45].

2.1.3. Age-dependent GC data analyses

Data on the risk of GC is obtained from the SEER database (https://seer.cancer.gov/) [46]. Gene expression data used for estimating the age-dependent synthesis rate of each target growth factor are from GTEx (https://gtexportal.org/home/) [47], reprocessed by using UCSC Xena [47,48], as described in Ref. [49].

2.2. Methods

2.2.1. Gene-expression data analysis tools

The read-count data in the TCGA database or other similar datasets can be analyzed using the "TCGAanalyze_DEA" function in the R package "TCGAbiollinks" [50].

The R package "limma" is used to analyze microarray gene-expression data, which supports the following three functions: (i) "lmFit" can fit linearly the probes for each gene for estimating the expression level of the gene; (ii) "eBayes" is used to adjust the linear model for improving the fitting accuracy, which is particularly suitable for datasets with small sample sizes or low variance; (iii) "topTable" can extract top-ranked samples, say, by *P*-values [51].

All the read-counts are converted to protein-TPM (pTPM) values [52]. Different datasets are normalized by the total expression level. For microarray datasets, the expression data are used directly.

The function "ComBat-seq" in the R package "sva" is applied to remove batch effects across different datasets under study.

A gene is considered as differentially expressed (DEG) if $\log_2 CPM > 0$ and $|\log_2 FC| \geq 0.5$ with *P*-value < 0.05, where *FC* is for fold change and *CPM* represents the count per million for the target gene's expression level.

The "corr.test" function in R package "psych" can calculate Pearson Correlation Coefficient (*PCC*) between genes in terms of their expressions. A pair of genes with $PCC > 0.3$ between their expression levels is considered as statistically correlated and as strongly statistically correlated if $PCC > 0.7$, both having *P*-value < 0.0001.

Z-score [53] of a value from a distribution measures the distance from the value to the mean value of the distribution in terms of the number of standard deviation units. The more distant, the more statistically significant the value is.

2.2.2. Pathway enrichment

Pathway enrichment analyses are conducted against pathways in three databases: KEGG (https://www.kegg.jp/), the oldest biological pathway database with pathway information collected from the published experimental studies and manually curated [54]; REACTOME (https://reactome.org/), which started as a molecular interaction database that has been gradually extended to a pathway database [55]; and the GO database (http://geneontology.org/), the currently most popular biological function database, offering cellular pathway, molecular function and subcellular compartment information for a specified gene list [56]. *P*-value < 0.0001 is used when selecting enriched pathways.

2.2.3. Estimation of SA accumulation rate on cancer cell surface

The expression levels of SA synthetase, sialyltransferase (ST), and sialidase genes are used to estimate the SA accumulation rate and the average level of electrostatic repulsion between an adjacent pair of cells (see Supplementary Data and Table S3).

The rate of SA placement onto cell surface is estimated via the following function:

$$R_{SAp} = \sqrt{CMAS \times \sum_{i=1}^{11} ST_i} \tag{2}$$

where *CMAS* is the expression for a gene for SA synthesis, and $\sum_{i=1}^{11} ST_i$ is the total expression of ST_i genes for transferring SA onto a glycan on cell surface, where ST_i refers to the i^{th} ST gene.

The rate of SA degradation is estimated as follows:

$$R_{SAd} = \sqrt{NEU1 \times CTSA} \tag{3}$$

where *NEU1* is the expression of the *NEU1* gene for degrading SAs and *CTSA* is the expression of the gene needed to form a complex with *NEU1* for SA degradation [40].

The following function is used to estimate the SA accumulation rate on the surface on cancer cells:

$$R_{SAa} = R_{SAp} - R_{SAd}/S_c \tag{4}$$

where S_c is the average surface area of a cell type, collected from the Bionumbers database [57].

The level of electrostatic repulsion between two adjacent cells can be calculated as follows, using the Coulomb's law: $k \times q_1 \times q_2/r^2$, where q_1 and q_2 are the charges on the two cells, which is each approximated using $(R_{SAp} - R_{SAd})$. Assuming that the cells of the same cancer type have approximately the same size, we can use the following to approximate the level of electrostatic repulsion between two adjacent cells:

$$\frac{(R_{SAp} - R_{SAd})^2}{r_c^2} \tag{5}$$

where r_c denotes the average inter-cell distance among neighboring cells in cancer type *c*.

2.2.4. Linear regression

Function "regsubsets" in the R package "leap" is used to construct a linear model based on the expression levels of specified genes, which iteratively considers all possible subsets of the predictive variables and evaluates each subset with a set of predefined model selection criteria. The goal is to find the best model that captures the relationship between predictive variables and response variables, while avoiding overfitting. The quality of a regression model can be assessed using adjusted-R^2.

The R^2 value can be adjusted by the number of variables used by the model. It can measure models with different numbers of predictors to determine the quality of a model fit. The adjusted-R^2 is calculated as

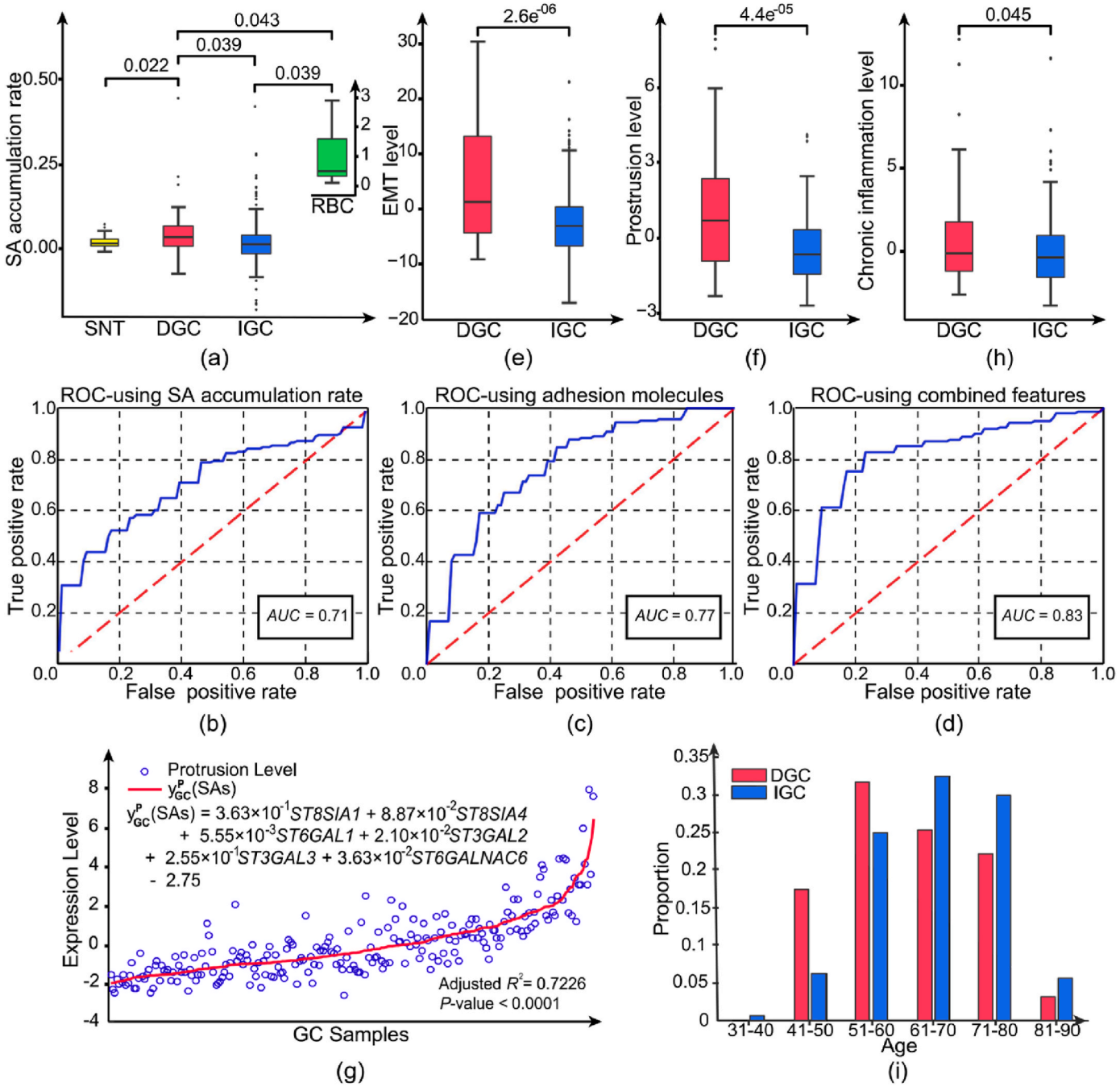

Fig. 1. Key analysis results for DGC and IGC samples. (a) The predicted SA accumulation rates in DGC, IGC, SNT and RBC samples, respectively. (b) The ROC of the classification results, using SA accumulation rates as the input feature. (c) The ROC of the classification results, using the expression of adhesion molecules as the input feature. (d) The ROC of the classification outcome, utilizing the combined features from (b) and (c). (e) The estimated EMT levels in DGCs and IGCs, respectively. (f) Assessing the protrusion levels of DGCs and IGCs. (g) Model fitting for protrusion levels using blue circles to represent GC protrusion levels based on the expressions of the relevant marker genes; a red curve to show predicted protrusion level $y^{P}_{GC}(SAs)$; and SA-related gene names to indicate their expression levels. (h) The estimated levels of chronic inflammation in DGCs and IGCs. (i) The age distributions of DGC and IGC samples.

follows:

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \tag{6}$$

$$\text{adjusted-}R^2 = 1 - \frac{(1-R^2)\bullet(n-1)}{(n-p-1)} \tag{7}$$

where SS_{res} represents the sum of residuals and SS_{tot} represents the total sum of squares; and n is the number of samples and p the number of variables.

2.2.5. Binary classification

The "SVC" function in Python is a classification method based on a Support Vector Machine (SVM). An SVM model is trained to identify the optimal hyperplane that maximizes the margin between the two classes of samples [58]. We have used a Gaussian kernel function in our application, defined as:

$$K(x, x^{'}) = \exp\left(-\gamma\|x - x^{'}\|^2\right) \tag{8}$$

where γ is a positive parameter to be determined through training, and $\|x - x'\|^2$ denotes the square of Euclidean distance.

A five-fold cross-validation is employed for verification of the classification. The receiver operating characteristic (ROC) curves and the area under the curve (AUC) are utilized to evaluate the classification performance of the SVM model.

2.2.6. Identification of the growth factor(s) for a specified cancer (sub)type

The following procedure is used to identify the growth factors specifically required by a specified set of cancer samples: (i) identify all upregulated growth-factor receptors in the given cancer samples vs. controls; and (ii) among the upregulated ones, identify all those strongly correlated with the cell-cycle genes. Using this procedure [49], we have identified the growth factors required by DGC and IGC, respectively.

2.2.7. Immune cell infiltration

Function "EPIC" in the R package "EPIC" (Estimation of Perturbations in Immune Cells) [59] has been utilized to estimate the levels of different immune cell types in a cancer sample. Seven immune cell types are considered in this study: B cells, CAFs (cancer-associated fibroblasts), $CD4^+$ T cells, $CD8^+$ T cells, endothelial cells, macrophages, and NK cells.

3. Results

3.1. Comparative analyses on DGCs and IGCs

3.1.1. Increased SAs on cell surface leads to stronger electrostatic repulsion between cells

Increased production and deployment of SAs on cancer cell surface have long been known to be associated with reduced cell-cell adhesion, increased immune activities, and cancer metastasis [39,40]. One distinguishing feature of SAs is their negative charge, resulting in electrostatic repulsion between adjacent cells, like in the case of red blood cells (RBCs) [40,43,44]. It is well-established that RBCs utilize the highest level of SAs among all human cell types, which prevents red blood cells from their aggregation [43].

We analyzed genes related to the deployment and degradation of SAs [41] (Table S1). The results show that five ST genes are expressed at higher levels in DGCs than in IGCs: *ST8SIA1*, *ST8SIA4*, *ST3GAL3*, *ST6GALNAC5*, and *ST6GALNAC6* (Table S2) while all other ST genes are expressed at comparable levels between the two subtypes. In addition, the main SA degradation enzyme, *NEU1* (sialidase) is downregulated by 0.61-fold in DGCs vs. IGCs. Collectively, these findings suggest that DGCs exhibit a higher density of SAs on their cell surface compared to IGCs.

We have estimated and compared the SA accumulation rates, denoted as R_{SAa}, of the DGC, IGC cells, and RBCs, using the method given in Methods (also Supplementary Data). The results reveal that the medians of R_{SAa} for DGC, IGC, SNT, and RBC samples are 0.0262, 0.0107, 0.0089, and 0.5057, respectively (Fig. 1(a)), revealing that DGCs have higher R_{SAa} values than IGCs and SNTs, but lower values than those of RBCs (P-value < 0.05). Knowing that the electrostatic repulsion between two adjacent cells is proportional to the product of the total surface charge of each cell, we conclude that the level of cell-cell repulsion between DGC cells is higher than four times of that between IGC cells with the same distance, but lower than that between RBCs. This provides a natural explanation of why DGC cells appear as small clusters rather than one large tumor.

To further demonstrate that stronger cell-cell repulsion in DGCs indeed plays a key role in formation of the tiny tumors compared to larger tumors of IGCs, a classification analysis is performed (see Methods). This analysis utilized the R_{SAa} values of DGCs and IGCs, in conjunction with the differential expression levels of various cell-cell adhesion genes (Table S3).

The classification results are shown in Fig. 1(b) - 1(d), showing that the AUC values by using the R_{SAa} values only, the expressions of the adhesion genes only, and using them both, are 0.71, 0.77, and 0.83, respectively. These results indicate that the estimated repulsion levels are highly consistent with the expressions of the adhesion genes, hence providing an independent support to our model.

3.1.2. Increased SA accumulation can lead to activation of EMT

EMT (Epithelial-Mesenchymal Transition) is a process that transforms an epithelial cell to a mesenchymal cell [60], known to be utilized in tissue development as well as in cancer migration and metastasis. The program, once activated, drives fundamental changes of an epithelial cell, including the loss of cell-cell adhesion, disruption of apical-basal polarity, and the development of migratory and invasive properties [61].

We use the expressions of 265 epithelial genes and 229 mesenchymal genes given in Ref. [62], to assess the epithelial level and mesenchymal level, represented as $Level_e$ and $Level_m$, respectively. We use ($Level_m - Level_e$) to measure how the host cells lean towards mesenchymal or epithelial trait in both DGC and IGC samples (Supplementary Data). The result is shown in Fig. 1(e), which reveals that DGCs have significantly higher mesenchymal traits than IGCs.

Elevation in the cell-surface sialylation level has been implicated in EMT activation [63]. To assess whether the upregulated ST genes are functionally related to the increased EMT level in DGCs vs. IGCs, we have examined the statistical correlation between the EMT levels and expressions of the ST genes across all 226 GC samples. We note that all the five upregulated ST genes correlate with the EMT level across all the GC samples (Table S4). Particularly, the *PCCs* between EMT vs. *ST8SIA1*, *ST3GAL3*, *ST6GALNAC5*, and *ST6GALNAC6* are all higher than 0.6, respectively.

Previous studies have established that persistent mechanical compression on cancer cells can lead to considerable cell deformation, which induces cellular contraction, protrusion, and the activation of the EMT program [40,64,65]. Based on this and the data presented in this section, we predict that the higher levels of production and deployment of SAs on DGC cell surface will lead to stronger cell-cell repulsion due to the higher levels of negative charges, and higher level of activation of the EMT program. Table S4 lists the supporting evidence that most of the mesenchymal marker genes co-expressed with STs are closely linked with cytoskeletal changes, cell deformation, and loss of cell polarity.

3.1.3. Enhanced SA accumulation may result in stronger migration

DGCs are known to have lower survival rates compared to IGCs [16]. It has been well established that over 93% of cancer-related death is due to metastasis across all cancers [66]. Hence, a higher death rate indicates a higher metastasis rate. We have checked if the over-accumulation of SAs is functionally linked to poorer survival rates of DGCs. To do this, we have assessed the level of cell migration by DGCs vs. that by IGCs. Here, we have assessed the level of cell protrusion, a key physical step in cell migration, the first step in cancer metastasis (Supplementary Data) [40]. Fig. 1(f) demonstrates that DGCs have a significantly higher protrusion level compared to IGCs.

A linear regression analysis is performed to assess the relationship between the protrusion level and expressions of the ST genes. We note that the heightened protrusion activity by DGCs can be adequately accounted for by the expressions of the ST genes as shown in Fig. 1(g). In the regression model, all the ST genes have a P-value < 0.05, reflecting their significant positive contributions to the increased cell protrusion. In addition, four out of the five upregulated ST genes appear in this model, namely *ST8SIA1*, *ST8SIA4*, *ST3GAL3*, and *ST6GALNAC6*, strongly supporting that the upregulated ST genes play major roles in the elevated level of protrusion in DGCs vs. IGCs.

We have also conducted co-expression analyses between each of the five upregulated ST genes and all DEGs across DGCs vs. IGCs, followed by a pathway enrichment analysis over these ST genes and the co-

Table 2
The number of the co-expressed DEGs and enriched pathways of the five STs.

	All pathways	Co-expressed DEGs	Developmental pathways	Immune-related pathways	Migration-related pathways
ST8SIA1	160	38	29 (18.13%)	20 (12.50%)	4 (2.50%)
ST8SIA4	517	150	37 (7.16%)	**191 (36.94%)**	24 (4.64%)
ST3GAL3	572	357	205 (35.84%)	3 (0.52%)	38 (6.64%)
ST6GALNAC5	524	232	215 (41.03%)	4 (0.76%)	46 (8.78%)
ST6GALNAC6	77	85	23 (29.87%)	1 (1.30%)	8 (10.39%)

expressed genes. All the pathways enriched by each group of co-expressed genes are given in Table S5, which fall largely into three functional categories: (i) migration-related pathways covering cell-cell adhesion, cell motility and migration; (ii) developmental pathways covering cell proliferation, development, differentiation, and morphogenesis; and (iii) immunoactivities. The percentage of the pathways in each of the three categories is shown in Table 2, revealing that the increased activities of these ST genes play important roles in migration-related activities, tumor development, and the induction of immunoactivities in DGC.

Knowing that integrins play essential roles in cancer migration, we have also examined the expressions of integrin genes in both DGCs and IGCs, and their *PCCs* with the ST genes (Table S6). We have observed the following. Four *ITGA* genes, *ITGA1*, *5*, *9*, and *V* are upregulated in DGC vs. IGC consistently across all four cancer stages except for *ITGA9* and *V* that have slightly higher expressions in stage IV IGC samples than the corresponding DGC samples. *ITGB1* has higher expression levels in DGC vs. IGC across all four stages; and *ITGB4* has lower expression levels in DGC vs. IGC across all four stages.

Very interestingly, these genes all have strong implications to survival. Specifically, upregulated *ITGA1*, *5*, *9*, *V*, and *ITGB1* all imply reduced survival rates for GC patients while upregulated *ITGB4* implies an improved survival rate of GC (Fig. S2). Furthermore, integrin pairs *ITGA2-ITGB1*, *ITGA6-ITGB1*, and *ITGA6-ITGB4* are known metastasis suppressors in various cancers [67–69], and all are downregulated in DGC vs. IGCs in at least three of the four stages. In addition, all the upregulated integrins are strongly correlated with at least three of the ST genes, which are upregulated in DGCs vs. IGCs (Table S6).

We have collected multiple supporting evidence from the literature that the upregulated ST genes tend to be associated with enhanced metastasis for multiple cancer types, and most of them are imvolved in the EMT activation and regulation of adhesion genes (Table S7). In addition, our survival analyses have shown that over-expressions of *ST8SIA4* and *ST6GALNAC6* have strong implications to reduced survival rates among all GC patients (P-value < 0.05) while increased expression of *ST3GAL3* has moderate implication to the reduced survival (P-value $= 0.12$) using the UALCAN server (Fig. S2).

3.1.4. Immune responses associated with ST genes in DGC vs. IGC

We have examined the expressions of a marker gene set for chronic inflammation across all the 226 GC samples, including chemokines, interleukins, and Toll-like receptors (Table S8); and observed considerably higher expressions of these genes in DGC than in IGC (P-value $<$ 0.05, Fig. 1(h)). We have then analyzed the level of immune infiltration by seven major types of immune cells and found that DGCs have higher infiltration levels of B cells, CAFs, $CD4^+$ T cells, endothelial cells, and macrophages than IGCs (P-value < 0.05, Fig. S3). None of the major immune cell types show higher infiltration levels in IGC than in DGC. Overall, DGC has higher inflammation level and higher immune infiltrations, which is consistent with published studies [32].

Poly-SAs (PSAs), synthesized by *ST8SIA4* and *ST8SIA2*, are generally used in neural development, predominantly in brain [70]. Hence, PSAs employed by GC cells will induce immune responses [71]. *ST8SIA4* generally produces longer PSA chains than *SIA8SIA2* [72]. We note that *ST8SIA4* is consistently upregulated in DGC than in IGC (Table S2). The largest functional category among pathways enriched by DEGs co-expressed with *ST8SIA4* is immunoactivities, accounting for 36.94% of the enriched pathways (Table 2). These pathways consist of the following activities: (i) innate immune activities such as chemotaxis, activation, differentiation and proliferation of neutrophils, leukocytes, natural killer (NK) cells, dendritic cell, mast cells and macrophages plus pathogen recognition signaling (e.g. Toll-like receptors signaling); (ii) T-cell activities: chemotaxis, TCR signaling, activation ($CD4^+$ T cells), selection, proliferation, differentiation (including helper T cells), apoptosis, lineage commitment, and cytokine production; and (iii) B-cell activities: BCR signaling, activation, differentiation proliferation and apoptosis, among a few others (Table S5).

The *PCCs* between *ST8SIA4* and the levels of chronic inflammation over all 226 GC samples were all at least 0.6355 (P-value < 0.0001), strongly suggesting that the level of inflammation is closely associated with the cell-surface accumulation of PSAs. In addition, we have also calculated the correlation between *ST8SIA4* and the level of infiltration by the seven immune cell types in GC samples. Table S8 shows the results for the infiltration levels by B cells, $CD4^+$ T cells, endothelial cells, and macrophages, with the macrophages having the highest *PCC* at 0.5883.

To further elucidate the detailed relationships between *ST8SIA4* and inflammatory cytokines in GC samples, we have examined the *PCC* levels between the expressions of *ST8SIA4* and interleukin (*IL*) genes. We note that six *IL* genes strongly correlate with *ST8SIA4*, namely *IL16*, *IL21*, *IL12B*, *IL15*, *IL34*, and *IL2* (Table S8), all being pro-inflammatory. They participate in both innate and adaptive immune activities such as differentiation, proliferation, and activities of macrophages (*IL21*, *IL34*), NK cells (*IL21*, *IL12B*, *IL15*), B cells (*IL21*, *IL2*), and T cells (*IL16*, *IL21*, *IL15*, *IL2*) [73]. It is noteworthy that *IL16* is the only upregulated *IL* in DGC samples (vs. IGC), a ligand of $CD4^+$ T cell. Taken these together, our results have revealed strong functional connections between the *ST8SIA4* expression and chronic inflammation levels, and possible modulations of immune responses by *ST8SIA4* in DGC.

Sialic acid-binding immunoglobulin-type lectins (*SIGLECs*) are the main SA receptors and known to play key regulatory roles in both innate and adaptive immunity [74]. We have calculated correlations between the expressions of *ST8SIA4* and each of all the 15 *SIGLEC* genes over all 226 GC samples; and discovered that *SIGLEC-1*, *2* (*CD22*), *3* (*CD33*), *7*, *8*, *9*, and *10* are upregulated in DGC vs. IGC, and each has a statistically significant correlation with *ST8SIA4* (Table S8). *SIGLEC-1* (*CD169*) is known to drive phagocytic functions in macrophages as well as stimulate pro-inflammatory cytokine production and lymphocyte proliferation [75,76]. *SIGLEC-2* (*CD22*) is an inhibitory receptor of BCR signaling [77]. *SIGLEC-3*, *7*, *8*, *9*, and *10* are mostly employed by innate immune system cells, and generally anti-inflammatory except for *SIGLEC-3*, *7*, which could each have both pro- and anti-inflammatory roles under specific conditions [78–80].

3.1.5. Younger DGC patients can be explained by specific growth factors

DGCs tend to occur in younger patients (age < 60) compared to IGC [16], but the reason remains elusive. We have estimated the age distributions of the patients from which the DGC and IGC samples are used here (Fig. 1(i)). It can be seen that a larger fraction of the DGC patients is younger than 60 years old compared to that of IGCs (P-value $= 0.0255$), which results in the total younger patient population.

Our previous research [49] has shown that (i) two factors, the

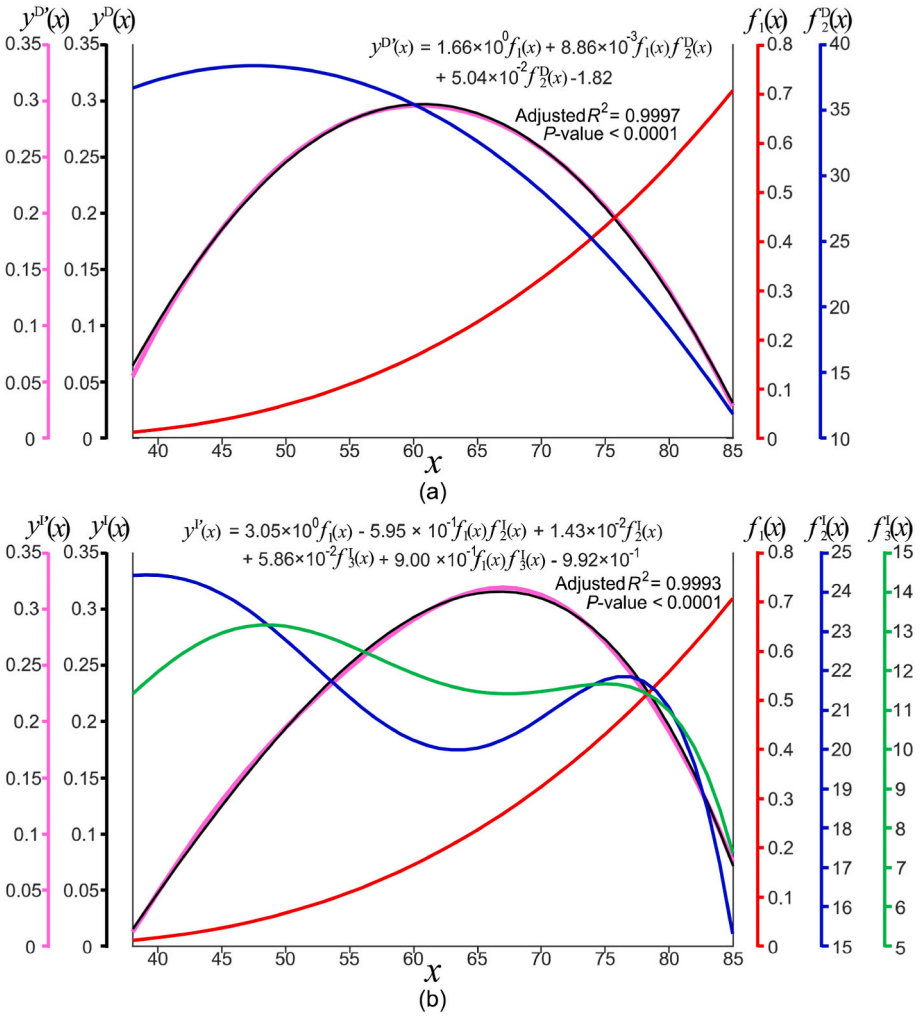

Fig. 2. Model fitting for DGC and IGC age-dependent occurrence rates. (a) Model fitting for DGC occurrence rates. The *x*-axis is for age. The purple curve represents the predicted DGC incidence ($y^{D'}(x)$). The black curve represents the known occurrence rate of DGC ($y^{D}(x)$). The red curve is for the cancer risk of stomach $f_1(x)$. The blue curve for the circulatory availability level of *PDGFC* ($f_2^{D}(x)$). (b) Model fitting for IGC occurrence rates. The *x*-axis is for age. The purple curve represents the predicted IGC incidence ($y^{I}(x)$). The black curve is for the known occurrence rate of IGC ($y^{I}(x)$). The red curve for the cancer risk of stomach $f_1(x)$. The blue and green curves are for the availability levels of *EREG* ($f_2^{I}(x)$) and *NRG2* ($f_3^{I}(x)$), respectively.

age-dependent risk of cancer development in the relevant organ and the age-dependent levels of the circulating growth factors specifically required by the cancer (sub)type determine the age-dependent occurrence rate of the cancer; and (ii) individuals who already have cancer-driving Fenton Reactions but lack the necessary levels of the growth factors required by the cancer (sub)type are more likely to experience tissue atrophy rather than cancer formation [81,82].

We have applied the model [49] to predict the growth factors specifically required by DGCs and IGCs, respectively. Two pieces of information are needed: age-dependent risk level for cancer development in an organ, stomach here [83]; and the growth factors needed by DGCs and by IGCs, respectively. The former is approximated by the risk of GC given in the SEER database [46]. And for the latter we have identified the upregulated growth-factor receptors, whose expressions strongly correlate with majority of the cell-cycle genes in DGCs and IGCs, respectively (see Methods). For DGCs and IGCs, we preliminarily predicted the respective growth factors of these two cancers, with their names and their receptors given in Tables S2B and S2C, respectively.

We have then performed a regression analysis of the age-dependent occurrence rate of each GC subtype against (i) the age-dependent risk level of cancer occurrence and (ii) the age-dependent blood concentrations of the predicted growth factors (see Supplementary Data). The result for DGCs is shown in Fig. 2(a), where the age-dependent occurrence rate $y^{D}(x)$ can be well-explained statistically by the known age-dependent risk level, $f_1(x)$, for cancer in stomach (collected from Ref. [49]) and the age-dependent level, $f_2^{D}(x)$, of the growth factor specifically required by DGCs, x for age, where *PDGFC* has been selected as a more representative key growth factor [84]. The model fitting achieves an accuracy level of $R^2 = 0.9997$ with P-value < 0.0001, indicating that the age-dependent risk curve and the age-dependent level of availability of the growth factor.

Similarly, *EREG* [85] and *NRG2* [86,87] are predicted to be the growth factors specifically required by IGCs. The occurrence rate $y^{I}(x)$ can be well-explained statistically by $f_1(x)$ and the age-dependent levels of *EREG* and *NRG2*, denoted by $f_2^{I}(x)$ and $f_3^{I}(x)$, respectively, as shown in Fig. 2(b). These results are supported by our literature review regarding the functions of the three growth factors (Table S7). The model fitting achieves an accuracy level of $R^2 = 0.9993$ with P-value < 0.0001, indicating that the age-dependent risk curve and age-dependent levels of the two growth factors.

We have calculated the average levels for the predicted growth factors for DGC and IGC in two age groups: 40–59 and 60–80, denoted as $C_{40\text{-}59}$ and $C_{60\text{-}80}$, respectively. Define $D = (C_{40\text{-}59} - C_{60\text{-}80})/C_{40\text{-}59} \times 100\%$ as the level of drop for a given growth factor in DGC or IGC. We

Table 3
Average concentrations and levels of drop for growth factors *PDGFC*, *EREG*, and *NRG2* between two age groups.

	$C_{40\text{-}59}$	$C_{60\text{-}80}$	D (%)
PDGFC	37.7087	28.1593	25.32
EREG	22.7194	20.8435	8.26
NRG2	12.7244	11.5445	9.27

note that *PDGFC* drops by $D = 25.32\%$ while the levels of the two growth factors for IGC drop less than 10% (Table 3). This means that the growth factor for DGC drops substantially from the first to the second age group while IGC does not have this behavior, hence providing a natural explanation to the observation.

We have previously established that (i) the rate of cancer proliferation is driven by the *de novo* biosynthesis rate of nucleotides, dictated by the OH^- production rate by the cytosolic Fenton reaction, which is ultimately due to chronic inflammation and local iron overload; and (ii) when there are no sufficient levels of growth factors needed specifically by the cancer, that matches the dictated proliferation rate, the cells will die [49], suggesting that tissue loss should be observed for such cancers. This is supported by the published report that DGC tends to be associated with atrophic gastritis [88].

3.2. Comparative analyses on other five cancer types

In a similar manner to the above, we have performed comparative analyses on four other cancer types, namely BC, PC, HC, and LC, each having a relatively large number of diffuse-like tumor samples with transcriptomic data available in the public domain (see Table S1). For these tumors, we focus on the following two questions rather than four as for GC since high-quality data for the other two questions are not available for these cancer types: (i) Do the diffuse-like tumors have higher SA accumulation rates than the corresponding non-diffuse subtypes? And (ii) Can the higher SA accumulation rate explain the higher metastatic potential and poorer prognosis of their diffuse-like tumors than the corresponding non-diffuse ones for each cancer type?

3.2.1. Higher SA accumulation rates in diffuse-like vs. non-diffuse tumors

We have estimated the SA accumulation rates for BC, PC and LC, respectively. IBC and non-IBC (NIBC) are the diffuse-like and non-diffuse subtypes of BC, respectively. For PC, a tumor with Gleason score ≥ 8 and Gleason grade ≥ 4 is considered as diffuse-like (DPC) while a tumor with Gleason score ≤ 6 and Gleason grade ≤ 3 is considered as non-diffuse subtype (NDPC). For LC, SCLC, and non-SCLC (NSCLC) are considered as diffuse-like and non-diffuse subtypes, respectively (Table 1).

Our analyses have revealed that for each of BC, PC, and LC, its diffuse-like tumors have higher levels of ST gene expressions and comparable expressions of the SA sialidase gene vs. the non-diffuse tumors

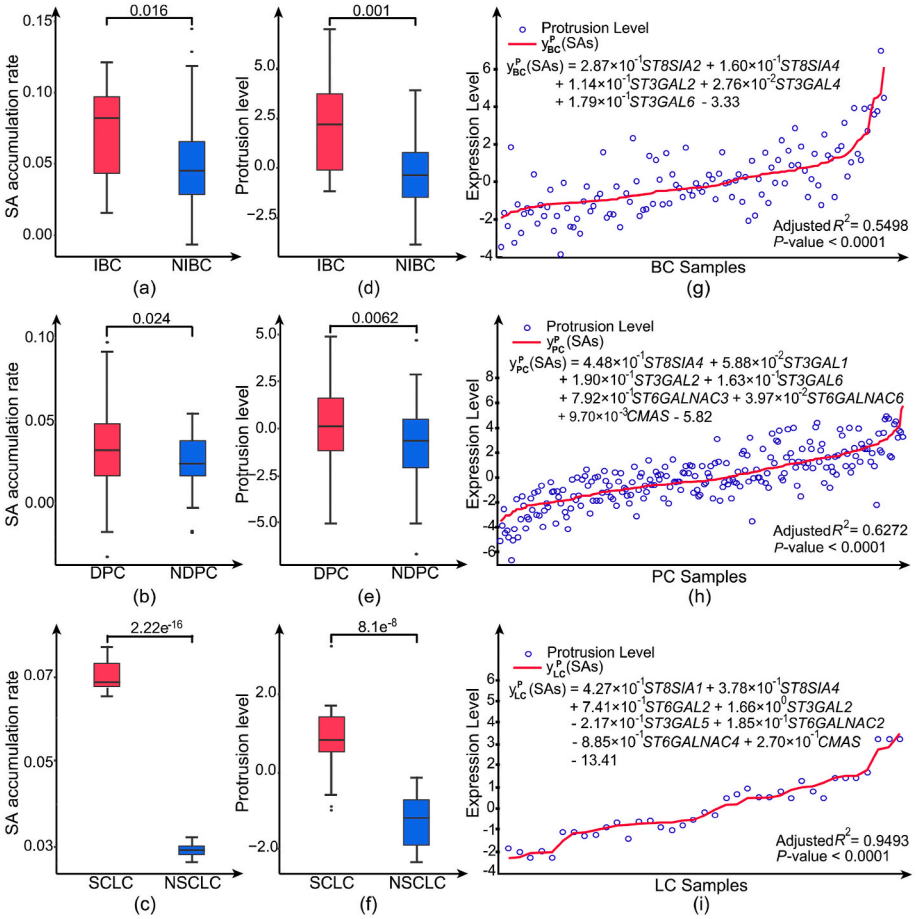

Fig. 3. Analysis results for BC, PC, and LC samples, respectively. (a) The predicted SA accumulation rates in IBC vs. NIBC samples, respectively. (b) The predicted SA accumulation rates in DPC and NDPC samples, respectively. (c) The predicted SA accumulation rates in SCLC vs. NSCLC samples, respectively. (d) The estimated protrusion levels in IBCs vs. NIBCs, respectively. (e) The estimated protrusion levels in DPCs vs. NDPCs. (f) The estimated protrusion levels in SCLCs vs. NSCLCs. (g) Model fitting for protrusion levels on BCs. The blue circles represent the protrusion level of BC reflected by the expression of the protrusion-related genes; the red curve for the predicted protrusion level $y^P_{BC}(SAs)$; and gene names denote their respective expression levels. BC samples were listed in the ascending order of their $y^P_{BC}(SAs)$ values from left to right along the *x*-axis. Similar is defined and performed for PC and LC samples in (h) and (i), respectively.

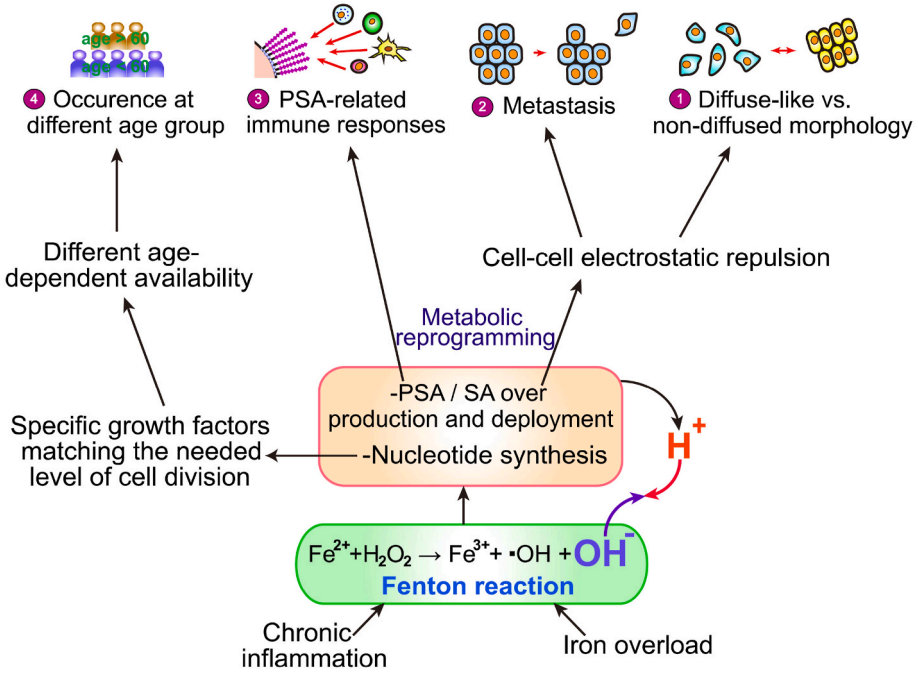

Fig. 4. The overall framework used for this study. The key points related to the four questions mentioned in the ABSTRACT are marked in the upper part.

(Table S2). The comparative R_{SAa} values for IBC and NIBC, DPC and NDPC, SCLC and NSCLC samples are shown in Fig. 3(a) - 3(c). The median R_{SAa} values for IBC and NIBC samples are 0.0820 and 0.0450, respectively. Those for DPC and NDPC are 0.0326 and 0.0246, respectively. And those for SCLC and NSCLC are 0.0688 and 0.0292, respectively. Therefore, the level of cell-cell repulsion in IBC is three times more than that in NIBC of the same distance in BC; DPC is 70% more than that in NDPC in PC; and SCLC is five times more than that in NSCLC in LC. Overall, the SA accumulation rate is consistently higher in diffuse-like vs. the matching non-diffuse tumors across all three cancer types. Hence, we predict that the higher cell-cell electrostatic repulsion is a key reason for the diffuse-like tumor subtypes.

3.2.2. Increased metastatic potential of diffuse-like vs. non-diffuse tumors

Analyses were conducted to assess the level of cell protrusion in diffuse-like and corresponding non-diffuse BC, PC, and LC tumors, respectively. We note that diffuse-like tumor samples have a significantly higher level of protrusion (P-value < 0.05) than non-diffuse tumors for each of the three cancer types, as detailed in Fig. 3(d) - 3(f). In addition, the increased protrusion activities can be well-explained statistically by the increased expressions of the SA synthesis gene (*CMAS*) and the ST genes in the diffuse-like vs. non-diffuse tumors for each of the three cancer types (Fig. 3(g)–3(i) and Supplementary Data). In these, all the ST genes and gene *CMAS* have P-values < 0.05.

Based on the above results, we infer that the elevated SA deployment plays key roles in the enhanced migration activities, metastatic potentials, and lower survival rates of the diffuse-like tumors vs. their non-diffuse counterparts in all three cancer types, which is well supported by published studies (Table S7).

We have also examined two other cancer types with limited diffuse-like samples and transcriptomic data in TCGA, namely liver cancer (HC) and thyroid cancer (TC) (Table S1). For both cancer types, the expressions of the SA transferase genes are higher while the expressions of the SA degradation genes are lower in the diffuse-like than in the non-diffuse tumors (Table S9), indicating that the diffuse-like tumors have higher levels of SA accumulation on cancer cell surface than the non-diffuse tumors. Furthermore, for numerous cancer types including these two, increased ST gene expressions are associated with reduced survival (Fig. S2), hence providing further evidence that SA over-deployment is associated with diffuse morphology and poorer prognoses.

4. Discussion

Through comparative analyses and modeling of transcriptomic data of diffuse/diffuse-like vs. non-diffuse cancer tissues, we have provided answers to each of the four questions posted in the beginning of this paper, with each having strong support from published studies. Fig. 4 summarizes the overall framework of our study.

As discussed in our previous study [38], different cancer types may activate a distinct combination of reprogrammed metabolisms to produce H^+ s at a level comparable to the level of OH^- production due to persistent Fenton reactions [36]. The determining factors of which reprogrammed metabolisms are activated may be dictated by (i) the level of OH^- production; and (ii) the ease for each reprogrammed metabolism to be induced under the current cellular conditions.

The key contributions of the present work can be summarized as follows, which are all reported for the first time, to the best of our knowledge:

1. identified key molecular determinants for the phenotypic differences between DGCs and IGCs, including that the higher level of SA accumulation on cancer cell surface in DGCs is the key reason for their distinct morphologies, more aggregative behavior, the higher death rates, and younger patients than IGCs;
2. identified the higher level of utilization of PSAs as the key reason for the higher levels of immunoactivities observed in DGCs vs. IGCs;
3. similar discoveries are made between diffuse-like and non-diffuse tumors of other cancer types including breast, liver, lung, prostate, and thyroid cancer, respectively, hence suggesting that our discoveries are general about cancer;
4. the discoveries made here provide new targets for biomarkers for early detection as well as for new drugs, such as inhibiting deployment of SAs on cancer cell surface, for the diffuse subtype of these cancer types.

The main limitations of the current framework include: (i) the analyses are largely qualitative rather than quantitative, resulting in a

limited level of resolution in our findings. For example, our analyses have revealed that DGCs deploy more SAs than IGCs but could not inform how much more SAs are accumulated on cancer cell surfaces in the two subtypes of GC tissues, which hindered us from getting more useful information for more accurate prognostic predictions. In addition, our model needs direct experimental validation.

5. Conclusion and future work

By statistical analyses of transcriptomic data of cancer tissues and physics-based arguments, we proposed a model for how the accumulation of negatively charged SAs on cancer cell surface, as way to overcome intracellular alkalization, can give rise to highly distinct morphologies as well as clinical behaviors between two subtypes of a cancer, namely diffuse vs. non-diffuse tumors. A key lesson we have learned here is that it is probably not the mis-signaling, which is often considered as a fundamental cause of a cancer but rather the reprogrammed metabolisms dictated by specific stressors and the physical properties developed due to the altered metabolisms. At a practical level, the knowledge generated here could lead to novel targets for treating the diffuse cancers, which are largely inoperable and hence considered as terminal currently.

Future work is planned to develop a quantitative model based on the current qualitative one, coupled with detailed chemical reactions represented as Michaelis-Menten equations and physics-based calculation of electrostatic repulsion among adjacent cells.

CRediT authorship contribution statement

Xuan Li: Writing – original draft, Validation, Software, Methodology, Investigation, Formal analysis, Conceptualization, Data curation. **Dingyun Liu:** Visualization, Software, Resources, Validation. **Zhipeng Wu:** Validation, Visualization. **Ying Xu:** Writing – review & editing, Supervision, Project administration, Conceptualization.

Declaration of competing interest

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Acknowledgements

This research was supported by the National Natural Science Foundation of China (No. T2350010) and Key University Laboratory of Metabolism and Health of Guangdong, Southern University of Science and Technology, Shenzhen 518055, China (No. 2022KSYS007).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compbiomed.2024.108703.

References

[1] I.A. Jaiyesimi, A.U. Buzdar, G. Hortobagyi, Inflammatory breast cancer: a review, J. Clin. Oncol. 10 (1992) 1014–1024, https://doi.org/10.1200/JCO.1992.10.6.1014.

[2] M.G.K. Benesch, A. Mathieson, Epidemiology of signet ring cell Adenocarcinomas, Cancers 12 (2020), https://doi.org/10.3390/cancers12061544.

[3] B. Hu, N. El Hajj, S. Sittler, N. Lammert, R. Barnes, A. Meloni-Ehrig, Gastric cancer: classification, histology and application of molecular pathology, J. Gastrointest. Oncol. 3 (2012) 251–261, https://doi.org/10.3978/j.issn.2078-6891.2012.021.

[4] S. Jakate, A. Yabes, D. Giusto, B. Naini, C. Lassman, M.M. Yeh, L.D. Ferrell, Diffuse cirrhosis-like hepatocellular carcinoma: a clinically and radiographically undetected variant mimicking cirrhosis, Am. J. Surg. Pathol. 34 (2010) 935–941, https://doi.org/10.1097/PAS.0b013e3181ddf52f.

[5] W.D. Travis, Update on small cell carcinoma and its differentiation from squamous cell carcinoma and other non-small cell carcinomas, Mod. Pathol. 25 (Suppl 1) (2012) S18–S30, https://doi.org/10.1038/modpathol.2011.150.

[6] H.U. Ahmed, M. Arya, A. Freeman, M. Emberton, Do low-grade and low-volume prostate cancers bear the hallmarks of malignancy? Lancet Oncol. 13 (2012) e509–e517, https://doi.org/10.1016/S1470-2045(12)70388-1.

[7] N. Chereau, X. Giudicelli, F. Pattou, J.C. Lifante, F. Triponez, E. Mirallie, P. Goudet, L. Brunaud, C. Tresallet, F. Tissier, L. Leenhardt, S.T. du Montcel, F. Menegaux, Diffuse sclerosing variant of papillary thyroid carcinoma is associated with aggressive Histopathological features and a poor outcome: results of a large Multicentric study, J. Clin. Endocrinol. Metab. 101 (2016) 4603–4610, https://doi.org/10.1210/jc.2016-2341.

[8] Y.C. Chen, W.L. Fang, R.F. Wang, C.A. Liu, M.H. Yang, S.S. Lo, C.W. Wu, A.F. Li, Y. M. Shyr, K.H. Huang, Clinicopathological Variation of Lauren classification in gastric cancer, Pathol. Oncol. Res. 22 (2016) 197–202, https://doi.org/10.1007/s12253-015-9996-6.

[9] M. Cristofanilli, V. Valero, A.U. Buzdar, S.W. Kau, K.R. Broglio, A.M. Gonzalez-Angulo, N. Sneige, R. Islam, N.T. Ueno, T.A. Buchholz, S.E. Singletary, G. N. Hortobagyi, Inflammatory breast cancer (IBC) and patterns of recurrence: understanding the biology of a unique disease, Cancer 110 (2007) 1436–1444, https://doi.org/10.1002/cncr.22927.

[10] P.M. Pierorazio, P.C. Walsh, A.W. Partin, J.I. Epstein, Prognostic Gleason grade grouping: data based on the modified Gleason scoring system, BJU Int. 111 (2013) 753–760, https://doi.org/10.1111/j.1464-410X.2012.11611.x.

[11] A.F. Gazdar, P.A. Bunn, J.D. Minna, Small-cell lung cancer: what we know, what we need to know and the path forward, Nat. Rev. Cancer 17 (2017) 725–737, https://doi.org/10.1038/nrc.2017.87.

[12] D.J. Heineman, J.M. Daniels, W.H. Schreurs, Clinical staging of NSCLC: current evidence and implications for adjuvant chemotherapy, Ther. Adv. Med. Oncol. 9 (2017) 599–609, https://doi.org/10.1177/1758834017722746.

[13] L. Benvegnu, F. Noventa, E. Bernardinello, P. Pontisso, A. Gatta, A. Alberti, Evidence for an association between the aetiology of cirrhosis and pattern of hepatocellular carcinoma development, Gut 48 (2001) 110–115, https://doi.org/10.1136/gut.48.1.110.

[14] K.H. Al-Qahtani, M. Al Asiri, M. Tunio, N.J. Aljohani, Y. Bayoumi, W. AlShakwer, Diffuse sclerosing variant papillary thyroid carcinoma: clinicopathological and treatment outcome analysis of 44 cases, Kuwait Med. J. 47 (2015) 225–230. https://www.researchgate.net/publication/281442254.

[15] E.G. Jones, Familial gastric cancer, N. Z.Med. J. 63 (1964) 287–296. https://pubmed.ncbi.nlm.nih.gov/14158754/.

[16] P. Lauren, The two Histological main types of gastric carcinoma: diffuse and So-Called intestinal-type carcinoma. An Attempt at a Histo-clinical classification, Acta Pathol. Microbiol. Scand. 64 (1965) 31–49, https://doi.org/10.1111/apm.1965.64.1.31.

[17] A. Isosalo, S.I. Inkinen, T. Turunen, P.S. Ipatti, J. Reponen, M.T. Nieminen, Independent evaluation of a multi-view multi-task convolutional neural network breast cancer classification model using Finnish mammography screening data, Comput. Biol. Med. 161 (2023) 107023, https://doi.org/10.1016/j.compbiomed.2023.107023.

[18] K. Hu, W. Chen, Y. Sun, X. Hu, Q. Zhou, Z. Zheng, PPNet: Pyramid pooling based network for polyp segmentation, Comput. Biol. Med. 160 (2023) 107028, https://doi.org/10.1016/j.compbiomed.2023.107028.

[19] M.A. Shamshiri, A. Krzyzak, M. Kowal, J. Korbicz, Compatible-domain transfer learning for breast cancer classification with limited annotated data, Comput. Biol. Med. 154 (2023) 106575, https://doi.org/10.1016/j.compbiomed.2023.106575.

[20] E. Özbay, F.A. Özbay, F.S. Gharehchopogh, Peripheral blood Smear Images classification for Acute Lymphoblastic leukemia diagnosis with an improved convolutional neural network, Journal of Bionic Engineering (2023), https://doi.org/10.1007/s42235-023-00441-y.

[21] J. Hu, S. Lv, T. Zhou, H. Chen, L. Xiao, X. Huang, L. Wang, P. Wu, Identification of Pulmonary Hypertension Animal models using a new Evolutionary machine learning framework based on blood Routine Indicators, J. Bionic Eng. 20 (2023) 762–781, https://doi.org/10.1007/s42235-022-00292-z.

[22] Z. Wu, H. Liu, J. Xie, G. Xu, G. Li, C. Lu, An effective method for the protection of user health topic privacy for health information services, World Wide Web 26 (2023) 1–23, https://doi.org/10.1007/s11280-023-01208-5.

[23] Y. Dai, J. Wu, Y. Fan, J. Wang, J. Niu, F. Gu, S. Shen, MSEva: A Musculoskeletal Rehabilitation Evaluation System Based on EMG Signals, vol. 19, ACM Trans. Sen. Netw., 2022 6, https://doi.org/10.1145/3522739.

[24] Z. Shen, P. Cao, J. Yang, O.R. Zaiane, WS-LungNet: a two-stage weakly-supervised lung cancer detection and diagnosis network, Comput. Biol. Med. 154 (2023) 106587, https://doi.org/10.1016/j.compbiomed.2023.106587.

[25] D.J. van Uden, H.W. van Laarhoven, A.H. Westenberg, J.H. de Wilt, C.F. Blanken-Peeters, Inflammatory breast cancer: an overview, Crit. Rev. Oncol. Hematol. 93 (2015) 116–126, https://doi.org/10.1016/j.critrevonc.2014.09.003.

[26] I.D. Nagtegaal, R.D. Odze, D. Klimstra, V. Paradis, M. Rugge, P. Schirmacher, K. M. Washington, F. Carneiro, I.A. Cree, W.H.O.C.o.T.E. Board, The 2019 WHO classification of tumours of the digestive system, Histopathology 76 (2020) 182–188, https://doi.org/10.1111/his.13975.

[27] M.G. Raso, N. Bota-Rabassedas, Wistuba, II, pathology and classification of SCLC, Cancers 13 (2021), https://doi.org/10.3390/cancers13040820.

[28] E.F. Clayton, S. Malik, A. Bonnel, Y. Mu, K. Olthoff, A. Shaked, P.L. Abt, H. Peterman, K. Rajender Reddy, S. Ottmann, E.E. Furth, M.H. Levine, Liver transplantation and cirrhotomimetic hepatocellular carcinoma: classification and outcomes, Liver Transplant. 20 (2014) 765–774, https://doi.org/10.1002/lt.23876.

[29] A.R. Reynolds, A. Furlan, D.T. Fetzer, E. Sasatomi, A.A. Borhani, M.T. Heller, M. E. Tublin, Infiltrative hepatocellular carcinoma: what radiologists need to know, Radiographics 35 (2015) 371–386, https://doi.org/10.1148/rg.352140114.

[30] A.L. Vickery Jr., M.L. Carcangiu, J.V. Johannessen, M. Sobrinho-Simoes, Papillary carcinoma, Semin. Diagn. Pathol. 2 (1985) 90–100. https://pubmed.ncbi.nlm.nih.gov/3843693/.
[31] G. Calderillo-Ruiz, M.C. Diaz-Romero, B. Carbajal-Lopez, M. Herrera-Martinez, E. Ruiz-Garcia, A.M. Leon-Takahashi, H.N. Lopez-Basave, A. Meneses-Garcia, A. Herrera-Gomez, Latin American young patients with gastric adenocarcinoma: worst prognosis and outcomes, J. Gastrointest. Oncol. 14 (2023) 2018–2027, https://doi.org/10.21037/jgo-23-259.
[32] J. Ma, H. Shen, L. Kapesa, S. Zeng, Lauren classification and individualized chemotherapy in gastric cancer, Oncol. Lett. 11 (2016) 2959–2964, https://doi.org/10.3892/ol.2016.4337.
[33] N.Y. Chia, P. Tan, Molecular classification of gastric cancer, Ann. Oncol. 27 (2016) 763–769, https://doi.org/10.1093/annonc/mdw040.
[34] T. Matsuoka, M. Yashiro, Rho/ROCK signaling in motility and metastasis of gastric cancer, World J. Gastroenterol. 20 (2014) 13756–13766, https://doi.org/10.3748/wjg.v20.i38.13756.
[35] S.C. Oh, B.H. Sohn, J.H. Cheong, S.B. Kim, J.E. Lee, K.C. Park, S.H. Lee, J.L. Park, Y.Y. Park, H.S. Lee, H.J. Jang, E.S. Park, S.C. Kim, J. Heo, I.S. Chu, Y.J. Jang, Y. J. Mok, W. Jung, B.H. Kim, A. Kim, J.Y. Cho, J.Y. Lim, Y. Hayashi, S. Song, E. Elimova, J.S. Estralla, J.H. Lee, M.S. Bhutani, Y. Lu, W. Liu, J. Lee, W.K. Kang, S. Kim, S.H. Noh, G.B. Mills, S.Y. Kim, J.A. Ajani, J.S. Lee, Clinical and genomic landscape of gastric cancer with a mesenchymal phenotype, Nat. Commun. 9 (2018) 1777, https://doi.org/10.1038/s41467-018-04179-8.
[36] H. Sun, C. Zhang, S. Cao, T. Sheng, N. Dong, Y. Xu, Fenton reactions drive nucleotide and ATP syntheses in cancer, J. Mol. Cell Biol. 10 (2018) 448–459, https://doi.org/10.1093/jmcb/mjy039.
[37] Y. Zhou, W. Chang, X. Lu, J. Wang, C. Zhang, Y. Xu, Acid-base homeostasis and implications to the phenotypic behaviors of cancer, Dev. Reprod. Biol. (2022), https://doi.org/10.1016/j.gpb.2022.06.003.
[38] H. Sun, Y. Zhou, M.F. Skaro, Y. Wu, Z. Qu, F. Mao, S. Zhao, Y. Xu, Metabolic reprogramming in cancer is induced to Increase proton production, Cancer Res. 80 (2020) 1143–1155, https://doi.org/10.1158/0008-5472.CAN-19-3392.
[39] E. Rodrigues, M.S. Macauley, Hypersialylation in cancer: Modulation of inflammation and therapeutic Opportunities, Cancers 10 (2018), https://doi.org/10.3390/cancers10060207.
[40] H. Sun, Y. Zhou, H. Jiang, Y. Xu, Elucidation of functional roles of sialic acids in cancer migration, Front. Oncol. 10 (2020) 401, https://doi.org/10.3389/fonc.2020.00401.
[41] K. Tomczak, P. Czerwinska, M. Wiznerowicz, The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge, Contemp. Oncol. 19 (2015) A68–A77, https://doi.org/10.5114/wo.2014.47136.
[42] T. Barrett, S.E. Wilhite, P. Ledoux, C. Evangelista, I.F. Kim, M. Tomashevsky, K. A. Marshall, K.H. Phillippy, P.M. Sherman, M. Holko, A. Yefanov, H. Lee, N. Zhang, C.L. Robertson, N. Serova, S. Davis, A. Soboleva, NCBI GEO: archive for functional genomics data sets–update, Nucleic Acids Res. 41 (2013) D991–D995, https://doi.org/10.1093/nar/gks1193.
[43] K.M. Jan, S. Chien, Role of surface electric charge in red blood cell interactions, J. Gen. Physiol. 61 (1973) 638–654, https://doi.org/10.1085/jgp.61.5.638.
[44] T.R. Klei, S.M. Meinderts, T.K. van den Berg, R. van Bruggen, From the Cradle to the Grave: the role of macrophages in Erythropoiesis and Erythrophagocytosis, Front. Immunol. 8 (2017) 73, https://doi.org/10.3389/fimmu.2017.00073.
[45] D.S. Chandrashekar, B. Bashel, S.A.H. Balasubramanya, C.J. Creighton, I. Ponce-Rodriguez, B. Chakravarthi, S. Varambally, UALCAN: a Portal for Facilitating tumor Subgroup gene expression and survival analyses, Neoplasia 19 (2017) 649–658, https://doi.org/10.1016/j.neo.2017.05.002.
[46] H.S. Park, S. Lloyd, R.H. Decker, L.D. Wilson, J.B. Yu, Overview of the Surveillance, Epidemiology, and End Results database: evolution, data variables, and quality assurance, Curr. Probl. Cancer 36 (2012) 183–190, https://doi.org/10.1016/j.currproblcancer.2012.03.007.
[47] G.T. Consortium, The Genotype-tissue expression (GTEx) project, Nat. Genet. 45 (2013) 580–585, https://doi.org/10.1038/ng.2653.
[48] M.J. Goldman, B. Craft, M. Hastie, K. Repecka, F. McDade, A. Kamath, A. Banerjee, Y. Luo, D. Rogers, A.N. Brooks, J. Zhu, D. Haussler, Visualizing and interpreting cancer genomics data via the Xena platform, Nat. Biotechnol. 38 (2020) 675–678, https://doi.org/10.1038/s41587-020-0546-8.
[49] S. Qiu, Z. An, R. Tan, P.A. He, J. Jing, H. Li, S. Wu, Y. Xu, Understanding the unimodal distributions of cancer occurrence rates: it takes two factors for a cancer to occur, Briefings Bioinf. 22 (2021), https://doi.org/10.1093/bib/bbaa349.
[50] A. Colaprico, T.C. Silva, C. Olsen, L. Garofano, C. Cava, D. Garolini, T.S. Sabedot, T. M. Malta, S.M. Pagnotta, I. Castiglioni, M. Ceccarelli, G. Bontempi, H. Noushmehr, TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data, Nucleic Acids Res. 44 (2016) e71, https://doi.org/10.1093/nar/gkv1507.
[51] R. Gentleman, Bioinformatics and Computational Biology Solutions Using R and Bioconductor, Springer Science+Business Media, New York, 2005. https://archive.org/details/bioinformaticsco0000unse/mode/2up.
[52] H. Jin, C. Zhang, M. Zwahlen, K. von Feilitzen, M. Karlsson, M. Shi, M. Yuan, X. Song, X. Li, H. Yang, H. Turkez, L. Fagerberg, M. Uhlen, A. Mardinoglu, Systematic transcriptional analysis of human cell lines for gene expression landscape and tumor representation, Nat. Commun. 14 (2023) 5417, https://doi.org/10.1038/s41467-023-41132-w.
[53] E. Lee, H.Y. Chuang, J.W. Kim, T. Ideker, D. Lee, Inferring pathway activity toward precise disease classification, PLoS Comput. Biol. 4 (2008) e1000217, https://doi.org/10.1371/journal.pcbi.1000217.
[54] M. Kanehisa, S. Goto, KEGG: kyoto encyclopedia of genes and genomes, Nucleic Acids Res. 28 (2000) 27–30, https://doi.org/10.1093/nar/28.1.27.
[55] A. Fabregat, S. Jupe, L. Matthews, K. Sidiropoulos, M. Gillespie, P. Garapati, R. Haw, B. Jassal, F. Korninger, B. May, M. Milacic, C.D. Roca, K. Rothfels, C. Sevilla, V. Shamovsky, S. Shorser, T. Varusai, G. Viteri, J. Weiser, G. Wu, L. Stein, H. Hermjakob, P. D'Eustachio, The reactome pathway Knowledgebase, Nucleic Acids Res. 46 (2018) D649–D655, https://doi.org/10.1093/nar/gkx1132.
[56] M. Ashburner, C.A. Ball, J.A. Blake, D. Botstein, H. Butler, J.M. Cherry, A.P. Davis, K. Dolinski, S.S. Dwight, J.T. Eppig, M.A. Harris, D.P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J.C. Matese, J.E. Richardson, M. Ringwald, G.M. Rubin, G. Sherlock, Gene ontology: tool for the unification of biology. The Gene Ontology Consortium, Nat. Genet. 25 (2000) 25–29, https://doi.org/10.1038/75556.
[57] R. Milo, P. Jorgensen, U. Moran, G. Weber, M. Springer, BioNumbers–the database of key numbers in molecular and cell biology, Nucleic Acids Res. 38 (2010) D750–D753, https://doi.org/10.1093/nar/gkp889.
[58] M. Wang, Y. Liang, Z. Hu, S. Chen, B. Shi, A.A. Heidari, Q. Zhang, H. Chen, X. Chen, Lupus nephritis diagnosis using enhanced moth flame algorithm with support vector machines, Comput. Biol. Med. 145 (2022) 105435, https://doi.org/10.1016/j.compbiomed.2022.105435.
[59] J. Racle, K. de Jonge, P. Baumgaertner, D.E. Speiser, D. Gfeller, Simultaneous enumeration of cancer and immune cell types from bulk tumor gene expression data, Elife 6 (2017), https://doi.org/10.7554/eLife.26476.
[60] C. Cumin, Y.L. Huang, A. Everest-Dass, F. Jacob, Deciphering the importance of Glycosphingolipids on cellular and molecular Mechanisms associated with epithelial-to-mesenchymal transition in cancer, Biomolecules 11 (2021), https://doi.org/10.3390/biom11010062.
[61] Z. Peng, C.X. Wang, E.H. Fang, G.B. Wang, Q. Tong, Role of epithelial-mesenchymal transition in gastric cancer initiation and progression, World J. Gastroenterol. 20 (2014) 5403–5410, https://doi.org/10.3748/wjg.v20.i18.5403.
[62] S. Huang, M. Holzel, T. Knijnenburg, A. Schlicker, P. Roepman, U. McDermott, M. Garnett, W. Grernrum, C. Sun, A. Prahallad, F.H. Groenendijk, L. Mittempergher, W. Nijkamp, J. Neefjes, R. Salazar, P. Ten Dijke, H. Uramoto, F. Tanaka, R.L. Beijersbergen, L.F. Wessels, R. Bernards, MED12 controls the response to multiple cancer drugs through regulation of TGF-beta receptor signaling, Cell 151 (2012) 937–950, https://doi.org/10.1016/j.cell.2012.10.035.
[63] F. Li, J. Ding, Sialylation is involved in cell fate decision during development, reprogramming and cancer progression, Protein Cell 10 (2019) 550–565, https://doi.org/10.1007/s13238-018-0597-5.
[64] J.M. Tse, G. Cheng, J.A. Tyrrell, S.A. Wilcox-Adelman, Y. Boucher, R.K. Jain, L. L. Munn, Mechanical compression drives cancer cells toward invasive phenotype, Proc. Natl. Acad. Sci. U. S. A. 109 (2012) 911–916, https://doi.org/10.1073/pnas.1118910109.
[65] N.E. Cabrera-Benitez, M. Parotto, M. Post, B. Han, P.M. Spieth, W.E. Cheng, F. Valladares, J. Villar, M. Liu, M. Sato, H. Zhang, A.S. Slutsky, Mechanical stress induces lung fibrosis by epithelial-mesenchymal transition, Crit. Care Med. 40 (2012) 510–517, https://doi.org/10.1097/CCM.0b013e31822f09d7.
[66] T.N. Seyfried, L.C. Huysentruyt, On the origin of cancer metastasis, Crit. Rev. Oncog. 18 (2013) 43–73, https://doi.org/10.1615/critrevoncog.v18.i1-2.40.
[67] Y. Ishii, A. Ochiai, T. Yamada, S. Akimoto, K. Yanagihara, M. Kitajima, S. Hirohashi, Integrin alpha6beta4 as a suppressor and a predictive marker for peritoneal dissemination in human gastric cancer, Gastroenterology 118 (2000) 497–506, https://doi.org/10.1016/s0016-5085(00)70255-1.
[68] H. Shimizu, T. Seiki, M. Asada, K. Yoshimatsu, N. Koyama, Alpha6beta1 integrin induces proteasome-mediated cleavage of erbB2 in breast cancer cells, Oncogene 22 (2003) 831–839, https://doi.org/10.1038/sj.onc.1206203.
[69] N.E. Ramirez, Z. Zhang, A. Madamanchi, K.L. Boyd, L.D. O'Rear, A. Nashabi, Z. Li, W.D. Dupont, A. Zijlstra, M.M. Zutter, The alpha(2)beta(1) integrin is a metastasis suppressor in mouse models and human cancer, J. Clin. Invest. 121 (2011) 226–237, https://doi.org/10.1172/JCI42328.
[70] A. Varki, P. Gagneux, Multifarious roles of sialic acids in immunity, Ann. N. Y. Acad. Sci. 1253 (2012) 16–36, https://doi.org/10.1111/j.1749-6632.2012.06517.x.
[71] P.M. Drake, J.K. Nathan, C.M. Stock, P.V. Chang, M.O. Muench, D. Nakata, J. R. Reader, P. Gip, K.P. Golden, B. Weinhold, R. Gerardy-Schahn, F.A. Troy 2nd, C. R. Bertozzi, Polysialic acid, a glycan with highly restricted expression, is found on human and murine leukocytes and modulates immune responses, J. Immunol. 181 (2008) 6850–6858, https://doi.org/10.4049/jimmunol.181.10.6850.
[72] K. Angata, M. Suzuki, M. Fukuda, Differential and cooperative polysialylation of the neural cell adhesion molecule by two polysialyltransferases, PST and STX, J. Biol. Chem. 273 (1998) 28524–28532, https://doi.org/10.1074/jbc.273.43.28524.
[73] M. Uhlen, L. Fagerberg, B.M. Hallstrom, C. Lindskog, P. Oksvold, A. Mardinoglu, A. Sivertsson, C. Kampf, E. Sjostedt, A. Asplund, I. Olsson, K. Edlund, E. Lundberg, S. Navani, C.A. Szigyarto, J. Odeberg, D. Djureinovic, J.O. Takanen, S. Hober, T. Alm, P.H. Edqvist, H. Berling, H. Tegel, J. Mulder, J. Rockberg, P. Nilsson, J. M. Schwenk, M. Hamsten, K. von Feilitzen, M. Forsberg, L. Persson, F. Johansson, M. Zwahlen, G. von Heijne, J. Nielsen, F. Ponten, Proteomics. Tissue-based map of the human proteome, Science 347 (2015) 1260419, https://doi.org/10.1126/science.1260419.
[74] M.S. Macauley, P.R. Crocker, J.C. Paulson, Siglec-mediated regulation of immune cell function in disease, Nat. Rev. Immunol. 14 (2014) 653–666, https://doi.org/10.1038/nri3737.
[75] P.L. Delputte, H. Van Gorp, H.W. Favoreel, I. Hoebeke, I. Delrue, H. Dewerchin, F. Verdonck, B. Verhasselt, E. Cox, H.J. Nauwynck, Porcine sialoadhesin (CD169/Siglec-1) is an endocytic receptor that allows targeted delivery of toxins and antigens to macrophages, PLoS One 6 (2011) e16827, https://doi.org/10.1371/journal.pone.0016827.

[76] Y.S. Xiong, Y. Cheng, Q.S. Lin, A.L. Wu, J. Yu, C. Li, Y. Sun, R.Q. Zhong, L.J. Wu, Increased expression of Siglec-1 on peripheral blood monocytes and its role in mononuclear cell reactivity to autoantigen in rheumatoid arthritis, Rheumatology 53 (2014) 250–259, https://doi.org/10.1093/rheumatology/ket342.

[77] J. Muller, L. Nitschke, The role of CD22 and Siglec-G in B-cell tolerance and autoimmune disease, Nat. Rev. Rheumatol. 10 (2014) 422–428, https://doi.org/10.1038/nrrheum.2014.54.

[78] S. Varchetta, E. Brunetta, A. Roberto, J. Mikulak, K.L. Hudspeth, M.U. Mondelli, D. Mavilio, Engagement of Siglec-7 receptor induces a pro-inflammatory response selectively in monocytes, PLoS One 7 (2012) e45821, https://doi.org/10.1371/journal.pone.0045821.

[79] T. Avril, H. Attrill, J. Zhang, A. Raper, P.R. Crocker, Negative regulation of leucocyte functions by CD33-related siglecs, Biochem. Soc. Trans. 34 (2006) 1024–1027, https://doi.org/10.1042/BST0341024.

[80] P.R. Crocker, J.C. Paulson, A. Varki, Siglecs and their roles in the immune system, Nat. Rev. Immunol. 7 (2007) 255–266, https://doi.org/10.1038/nri2056.

[81] R.H. Verhoeven, H.E. Karim-Kos, J.W. Coebergh, M. Brink, S. Horenblas, R. de Wit, B.A. Kiemeney, Markedly increased incidence and improved survival of testicular cancer in The Netherlands, Acta Oncol. 53 (2014) 342–350, https://doi.org/10.3109/0284186X.2013.819992.

[82] H. Moller, A. Prener, N.E. Skakkebaek, Testicular cancer, cryptorchidism, inguinal hernia, testicular atrophy, and genital malformations: case-control studies in Denmark, Cancer Causes Control 7 (1996) 264–274, https://doi.org/10.1007/BF00051302.

[83] C. Tomasetti, B. Vogelstein, Cancer etiology. Variation in cancer risk among tissues can be explained by the number of stem cell divisions, Science 347 (2015) 78–81, https://doi.org/10.1126/science.1260825.

[84] H. Yoon, C.M. Tang, S. Banerjee, M. Yebra, S. Noh, A.M. Burgoyne, J. Torre, M. Siena, M. Liu, L.R. Klug, Y.Y. Choi, M. Hosseini, A.L. Delgado, Z. Wang, R. P. French, A. Lowy, R.P. DeMatteo, M.C. Heinrich, A.A. Molinolo, J.S. Gutkind, O. Harismendy, J.K. Sicklick, Cancer-associated fibroblast secretion of PDGFC promotes gastrointestinal stromal tumor growth and metastasis, Oncogene 40 (2021) 1957–1973, https://doi.org/10.1038/s41388-021-01685-w.

[85] C. Wang, Q. Long, Q. Fu, Q. Xu, D. Fu, Y. Li, L. Gao, J. Guo, X. Zhang, E.W. Lam, J. Campisi, Y. Sun, Targeting epiregulin in the treatment-damaged tumor microenvironment restrains therapeutic resistance, Oncogene 41 (2022) 4941–4959, https://doi.org/10.1038/s41388-022-02476-7.

[86] M.L. Slattery, E.M. John, M.C. Stern, J. Herrick, A. Lundgreen, A.R. Giuliano, L. Hines, K.B. Baumgartner, G. Torres-Mejia, R.K. Wolff, Associations with growth factor genes (FGF1, FGF2, PDGFB, FGFR2, NRG2, EGF, ERBB2) with breast cancer risk and survival: the Breast Cancer Health Disparities Study, Breast Cancer Res. Treat. 140 (2013) 587–601, https://doi.org/10.1007/s10549-013-2644-5.

[87] C. Xu, Q. Wang, D. Wang, W. Wang, W. Fang, Z. Li, A. Liu, J. Yu, W. Zhong, Z. Wang, Y. Zhang, J. Liu, S. Zhang, X. Cai, A. Liu, W. Li, P. Zhan, H. Liu, T. Lv, L. Miao, L. Min, Y. Chen, J. Yuan, F. Wang, Z. Jiang, G. Lin, L. Huang, X. Pu, R. Lin, W. Liu, C. Rao, D. Lv, Z. Yu, X. Li, C. Tang, C. Zhou, J. Zhang, J. Xue, H. Guo, Q. Chu, R. Meng, J. Wu, R. Zhang, J. Zhou, Z. Zhu, Y. Li, H. Qiu, F. Xia, Y. Lu, X. Chen, R. Ge, E. Dai, Y. Han, W. Pan, F. Pang, Q. He, J. Huang, K. Wang, F. Wu, B. Xu, L. Wang, Y. Zhu, L. Lin, Y. Xie, X. Lin, J. Cai, L. Xu, J. Li, X. Jiao, K. Li, J. Wei, H. Feng, L. Wang, Y. Du, W. Yao, X. Shi, X. Niu, D. Yuan, Y. Yao, J. Huang, Y. Feng, Y. Zhang, P. Sun, H. Wang, M. Ye, Z. Wang, Y. Hao, Z. Wang, B. Wan, D. Lv, S. Yang, J. Kang, J. Zhang, C. Zhang, J. Ou, L. Shi, Y. Wang, B. Li, Z. Zhang, Z. Li, Z. Liu, N. Yang, L. Wu, H. Wang, G. Jin, G. Wang, J. Wang, M. Fang, Y. Fang, Y. Li, X. Wang, Y. Zhang, X. Zhu, Y. Shen, S. Ma, B. Wang, L. Si, Y. Song, Y. Lu, J. Chen, Z. Song, Expert Consensus on the diagnosis and treatment of NRG1/2 gene Fusion solid tumors, Glob. Med. Genet. 11 (2024) 86–99, https://doi.org/10.1055/s-0044-1781457.

[88] S.Y. Shin, J.H. Kim, J. Chun, Y.H. Yoon, H. Park, Chronic atrophic gastritis and intestinal metaplasia surrounding diffuse-type gastric cancer: are they just bystanders in the process of carcinogenesis? PLoS One 14 (2019) e0226427 https://doi.org/10.1371/journal.pone.0226427.