

MOLECULAR BIOLOGY & GENETICS

Molecular bases of morphologically diffused tumors across multiple cancer types

Dingyun Liu ^{1,2}, Feiyang Xing ^{2,3}, Yueying Wang¹, Jun Xiao^{1,2}, Zheng An^{2,4} and Ying Xu^{2,4,*}

Diffuse gastric cancer (DGC) and intestinal gastric cancer (IGC) are the two main gastric cancer subtypes [1]. Morphologically, DGC consists of numerous disjoint tiny tumors while IGC forms one or a few solid tumors (Supplementary Supplementary Fig. S1) [1,2]. In-depth studies have discovered that DGCs generally have an ill-defined or missing glandular structure and reduced cell–cell adhesion, and may tend to consist of a large fraction of signet cells [3] compared to IGCs. They also tend to be more aggressive and happen to younger patients compared to IGCs [1]. A number of studies have been published regarding the genomic and transcriptomic differences between DGC and IGC [4,5]. However, no studies have established how these molecular-level differences are functionally linked to the distinct morphologies and aggressiveness between DGCs and IGCs.

We have discovered based on transcriptomic data analyses that DGCs generally accumulate considerably more sialic acids (SAs) on the cancer cell surfaces compared to IGCs, which will result in stronger cell–cell electrostatic repulsion since SAs each carry a negative charge and hence on average larger cell–cell distances similar to that among red blood cells. Statistical analyses reveal that there is a strong association between the elevated expression levels of the relevant genes and the reduced survival time. These lead to our main hypothesis: *the level of SA accumulation on the cancer cell surfaces is a key factor that dictates*

the morphology and the aggressiveness of DGCs vs. IGCs.

Literature search has revealed that multiple other cancer types also each consist of subtypes like diffuse vs. non-diffused tumors of gastric cancer, namely scattered tiny tumors with ill-defined glandular structures, summarized in Supplementary Supplementary Table S1. Our further analyses provided strong evidence that this hypothesis applies to these cancer types as well.

MORE SAs ON CELL SURFACES IMPLY STRONGER CELL–CELL REPULSION AND METASTATIC POTENTIAL

It has long been observed since the 1960s that overproduction of SAs, which are the capping molecules of cell-surface glycans, is associated with cancer metastasis in general [6]. Each SA carries a negative charge and hence its over-deployment on the cell surfaces will lead to stronger electrostatic repulsion among neighboring cells, similar to red blood cells (RBCs) [7], which are known to deploy significantly more SAs on their surface than other cell types, preventing them from aggregation [8]. Hence, we have examined the expression levels of genes relevant to the SA accumulation on the cell surfaces, namely the sialyltransferase (ST) genes for deploying SAs and sialidase genes for degrading SAs.

We note that DGC samples (Supplementary Table S2) generally have higher levels of ST gene expressions and

lower sialidase gene expressions, suggesting that DGCs have more SAs accumulated on their cell surfaces than IGCs (Supplementary Table S3). This is supported by a quantitative analysis using the Michaelis–Menten kinetic model [7] (Supplementary material), which suggests that the SA accumulation rate and hence the cell–cell repulsion in DGCs are significantly higher than those of IGCs (Fig. 1A and Supplementary material).

Knowing that DGCs have higher death rates compared to IGCs [1], we have then analysed the levels of cell protrusion: a key migration-related activity [7] (Supplementary material), since metastasis is the predominant reason for cancer death, accounting for >93% of cancer-related death. Our analyses show that the protrusion level is considerably higher in DGCs than in IGCs (Fig. 1E). A linear regression analysis confirms that the increased protrusion activity can be statistically well explained by the expressions of the ST genes in DGCs (Fig. 1I and Supplementary material). This is further supported by the survival data associated with integrin genes in DGCs and IGCs (Supplementary Table S4, Supplementary Fig. S2 and Supplementary material).

YOUNGER DGC PATIENTS ARE DUE TO THE SPECIFIC GROWTH FACTOR USED

Our previous study suggests that more malignant cancers tend to happen in younger patients [9]. Specifically, a

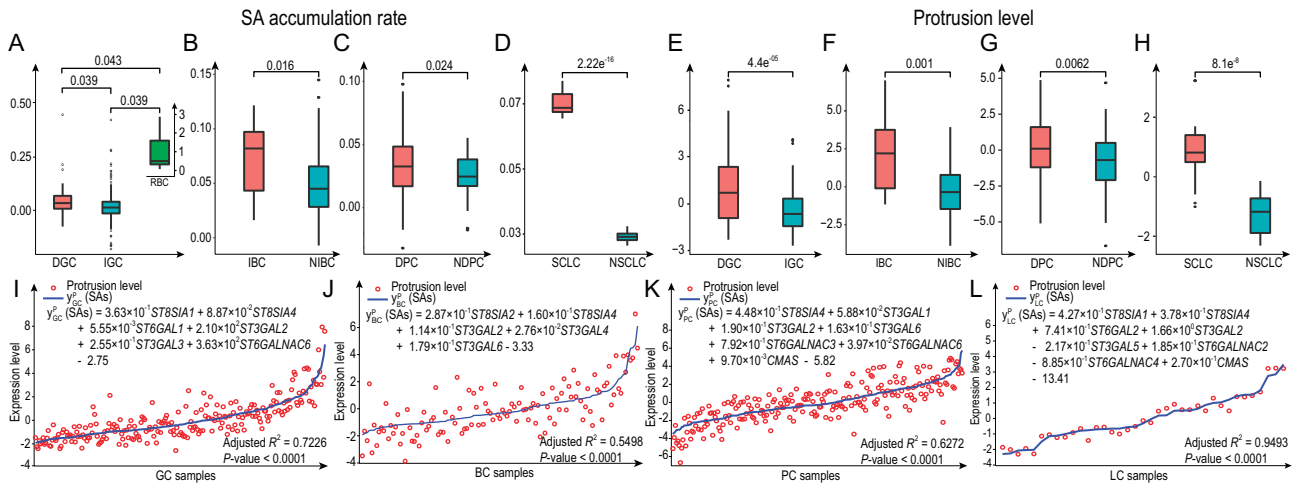


Figure 1. Key statistical results. (A) The predicted SA accumulation rates in DGC, IGC and RBC samples. (B) The predicted SA accumulation rates in IBC and NIBC samples. (C) The predicted SA accumulation rates in DPC and NDPC samples. (D) The predicted SA accumulation rates in SCLC and NSCLC samples. (E) The estimated protrusion levels in DGCs and IGCs. (F) The estimated protrusion levels in IBCs and NIBCs. (G) The estimated protrusion levels in DPCs and NDPCs. (H) The estimated protrusion levels in SCLCs and NSCLCs. (I) Model fitting for protrusion levels on GCs. The red circles represent the protrusion level of GC reflected by the expression of the protrusion-related genes; the blue curve represents the predicted protrusion level y_{GC}^P (SAs); and ST gene names denote their respective expression levels. GC samples were listed in the ascending order of their y_{GC}^P (SAs) values from left to right along the x-axis. For the corresponding density Distribution curves of protrusion levels and y_{GC}^P (SAs) values, x-axis denotes expression level and y-axis denotes sample proportion. The green curves $P_{diffuse}(x)$ denote the probability of the sample with protrusion level x [or y_{GC}^P (SAs) value x] being the diffuse-like sample, calculated from the corresponding density distribution. Similar is defined and performed for BC, PC and LC samples in (J), (K) and (L), respectively. Detailed methods are given in Supplementary material.

cancer requires two types of factors to take place, one being the cancer risk factor in an organ, which goes up with age, and the other the availability level of circulatory growth factors specifically needed by a cancer type, which decreases with age.

Using the program developed in [9], we have predicted the growth factors specifically needed by DGCs and IGCs, respectively (Supplementary material), with *PDGFC* being the main growth factor needed by DGCs and *EREG* and *NRG2* the growth factors needed by IGCs (Supplementary Table S3 and Supplementary material), which is supported by published studies (Supplementary Table S5) and by an accurate regression analysis (Supplementary Fig. S3 and Supplementary material). Similar is done for IGCs (Supplementary Fig. S3 and Supplementary material). Furthermore, the average level of growth factor *PDGFC* drops more sharply from the population with age <60 years to that with age >60 years than those of *EREG* and *NRG2*, hence providing a natural explanation for the observation (Supplementary material).

COMPARATIVE ANALYSES ON OTHER CANCER TYPES

Literature search has revealed that multiple other cancer types also each have tumor subtypes similar to DGCs in terms of their morphology and aggressiveness, as summarized in Supplementary Table S1. For each such cancer type, we call their DGC-like tumors diffuse-like tumors. Specifically, breast cancer (BC), prostate cancer (PC) and lung cancer (LC) each have a relatively large number of diffuse-like tumor samples with transcriptomic data in the public domain (Supplementary Table S2). In addition, other cancer types recorded in Supplementary Table S1 also have very limited diffuse-like tumor samples but we do not include them in the following analyses as their numbers of samples are too small.

SA deployment

The same analyses were conducted on BC, PC and LC. Inflammatory breast cancer (IBC), PC with Gleason score of ≥ 8 (DPC) and small-cell lung carcinoma (SCLC) are considered

to be diffuse-like subtypes of BC, PC and LC, respectively, while non-IBC (NIBC), PCs with Gleason score of ≤ 6 (NDPC) and non-SCLC (NSCLC) are the corresponding non-diffused subtypes (Supplementary Table S1). Our analyses have revealed that for each of the three cancer types, its diffuse-like tumors, on average, have higher levels of ST gene expressions (Supplementary Table S3). And Michaelis–Menten kinetics-based calculations show that the SA accumulation rates and cell–cell repulsion in the three diffuse-like tumors are consistently higher than those of non-diffused tumors (Fig. 1B–D and Supplementary material). Furthermore, the level of the predicted cell–cell repulsion is consistent with the survival rate of each diffuse-like subtype (Supplementary Table S1).

Increased metastatic potential

Analyses were then conducted to estimate the level of cell protrusion in the diffuse-like vs. non-diffused BC, PC and LC tumors, respectively. Comparable results are achieved here compared to those for DGCs vs. IGCs (Fig. 1F–H).

In addition, the increased protrusion activities in diffuse-like tumors can be statistically well explained by the expressions of the SA synthesis gene and ST genes for each of the three cancer types (Fig. 1J–L and Supplementary material). Based on these, we predict that the elevated SA deployment plays important roles in the increased migration activities and metastatic potentials of the diffuse-like subtypes of all three cancer types, which is further supported by published studies (Supplementary Table S5).

SUMMARY

Our data analyses and computational modeling provide support for the hypothesis that the distinct levels of SA accumulation on cancer cell surfaces are the key reason for the different morphologies and the aggressiveness between diffuse-like and non-diffused tumors of the same cancer types, hence providing new insights for an important and fundamental cancer biology question for the first time. The key novelty of our study lies in the fact that higher levels of the SA accumulation, estimated based on transcriptomic data, coupled with a physics-based argument provide a natural explanation for the possible causes of the distinct morphology as well as aggressiveness between diffuse-like vs. non-diffused tumors, plus the generality of this discovery. Clearly, this is a

computation-based discovery. Further validation is needed by physically measuring the levels of cell–cell repulsion in diffuse-like vs. non-diffused tumors and establishing the detailed relationship between the repulsion levels and the tumor sizes.

DATA AVAILABILITY

The data used in this work are openly and freely available in TCGA (<https://portal.gdc.cancer.gov/>), GEO (<https://www.ncbi.nlm.nih.gov/geo/>), SEER (<https://seer.cancer.gov/>), GTEx (<https://gtexportal.org/home/>), HUMAN PROTEIN ATLAS (<https://www.proteinatlas.org/>), Bionumbers (<https://bionumbers.hms.harvard.edu/>), and UALCAN (<http://ualcan.path.uab.edu/>).

SUPPLEMENTARY DATA

Supplementary data are available at *NSR* online.

ACKNOWLEDGEMENTS

The authors thank Georgia Research Alliance for continuous support to the corresponding author.

AUTHOR CONTRIBUTIONS

Dingyun Liu: design of the method, programming, data analysis, paper writing and revision. Feiyang Xing and Ying Xu: design of the methods, paper

writing and revision. Yueying Wang: programming and data processing. Jun Xiao and Zheng An: data collection and processing.

Conflict of interest statement. None declared.

Dingyun Liu ^{1,2}, Feiyang Xing ^{2,3}, Yueying Wang¹, Jun Xiao^{1,2}, Zheng An^{2,4} and Ying Xu^{2,4,*}

¹College of Computer Science and Technology, Jilin University, China; ²Center for Cancer Systems Biology, China-Japan Union Hospital of Jilin University, China; ³School of Life Sciences, Jilin University, China and ⁴Computational Systems Biology Lab, Department of Biochemistry and Molecular Biology and Institute of Bioinformatics, University of Georgia, USA

*Corresponding author. E-mail: xyn@uga.edu

REFERENCES

1. Lauren P. *Acta Pathol Microbiol Scand* 1965; **64**: 31–49.
2. Rumba R, Vanags A and Kalva A *et al. Surgical Management of Malignant Gastric Tumours: A Practical Guide*. London: InTech, 2017.
3. Benesch MGK and Mathieson A. *Cancers* 2020; **12**: 1544.
4. Chia NY and Tan P. *Ann Oncol* 2016; **27**: 763–9.
5. Oh SC, Sohn BH and Cheong JH *et al. Nat Commun* 2018; **9**: 1777.
6. Sun H, Zhou Y and Skaro MF *et al. Cancer Res* 2020; **80**: 1143–55.
7. Sun H, Zhou Y and Jiang H *et al. Front Oncol* 2020; **10**: 401.
8. Jan KM and Chien S. *J Gen Physiol* 1973; **61**: 638–54.
9. Qiu S, An Z and Tan R *et al. Brief Bioinform* 2020; **9**: bbaa349.