

Integrated Analysis Identifies Molecular Signatures and Specific Prognostic Factors for Different Gastric Cancer Subtypes¹



Li Min^{*,†,2}, Yu Zhao^{*,2}, Shengtao Zhu^{*},
Xintao Qiu[‡], Rui Cheng^{*}, Jie Xing^{*}, Linlin Shao^{*},
Shuilong Guo^{*} and Shutian Zhang^{*}

^{*}Department of Gastroenterology, Beijing Friendship Hospital, Capital Medical University, National Clinical Research Center for Digestive Disease, Beijing Digestive Disease Center, Beijing Key Laboratory for Precancerous Lesion of Digestive Disease, Beijing, 100050, PR China;

[†]Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, MA 02115, USA;

[‡]Department of Biomedical Informatics, Harvard School of Public Health, Boston, MA 02115, USA

Abstract

BACKGROUND: Gastric cancer (GC) is the fifth leading cause of cancer-related deaths worldwide. As an effective and easily performed method, microscopy-based Lauren classification has been widely accepted by gastrointestinal surgeons and pathologists for GC subtyping, but molecular characteristics of different Lauren subtypes were poorly revealed. **METHODS:** GSE62254 was used as a derivation cohort, and GSE15459 was used as a validation cohort. The difference between diffuse and intestinal GC on the gene expression level was measured. Gene ontology (GO) enrichment analysis was performed for both subgroups. Hierarchical clustering and heatmap exhibition were also performed. Kaplan-Meier plot and Cox proportional hazards model were used to evaluate survival grouped by the given genes or hierarchical clusters. **RESULTS:** A total of 4598 genes were found differentially expressed between diffuse and intestinal GC. Immunity- and cell adhesion-related GOs were enriched for diffuse GC, whereas DNA repair- and cell cycle-related GOs were enriched for intestinal GC. We proposed a 40-gene signature ($\chi^2 = 30.71, P < .001$) that exhibits better discrimination for prognosis than Lauren classification ($\chi^2 = 12.11, P = .002$). FRZB [RR (95% CI) = 1.824 (1.115-2.986), $P = .017$] and EFEMP1 [RR (95% CI) = 1.537 (0.969-2.437), $P = .067$] were identified as independent prognostic factors only in diffuse GC but not in intestinal GC patients. KRT23 [RR (95% CI) = 1.616 (0.938-2.785), $P = .083$] was identified as an independent prognostic factor only in intestinal GC patients but not in diffuse GC patients. Similar results were achieved in the validation cohort. **CONCLUSION:** We found that GCs with different Lauren classifications had different molecular characteristics and identified FRZB, EFEMP1, and KRT23 as subtype-specific prognostic factors for GC patients.

Translational Oncology (2017) 10, 99–107

Introduction

With estimated 951,600 new cases and 723,100 deaths in 2012 worldwide, gastric cancer (GC) still ranks fifth in incidence and third in mortality among all types of cancer [1]. The overall global incidence is declining during recent decades, especially in populations of high socioeconomic status, yet certain subtypes of GC showed a continuing increase even in developed countries such as United States [2]. Nearly 95% of GC cases are adenocarcinoma, and strong heterogeneity also exists among gastric adenocarcinoma cases [3]. The Lauren system is the most commonly used classification method which has been proven useful in evaluating the natural carcinogenesis

Address all correspondence to: Prof. Shutian Zhang, MD, PhD, or Shuilong Guo, PhD, Department of Gastroenterology, Beijing Friendship Hospital, Capital Medical University, National Clinical Research Center for Digestive Disease, Beijing Digestive Disease Center, Beijing Key Laboratory for Precancerous Lesion of Digestive Disease, Beijing, 100050, PR China.
E-mail: zhangshutian@ccmu.edu.cn, slong.guo@163.com

¹Funding: This study was supported by the National Key Technologies R&D Program (2015BAI13B09), National Natural Science Foundation of China (81302160, 81272447, 81670474), and Natural Science Foundation of Beijing (7152043).

²Li Min and Yu Zhao contributed equally to this work.

Received 5 September 2016; Revised 7 November 2016; Accepted 10 November 2016

© 2016 The Authors. Published by Elsevier Inc. on behalf of Neoplasia Press, Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).
1936-5233/17

<http://dx.doi.org/10.1016/j.tranon.2016.11.003>

history of GC patients [4,5]. Depending on the morphology, the Lauren system divides gastric adenocarcinoma into two distinct histological subtypes: diffuse GC and intestinal GC [4].

Diffuse GC could be characterized by scattered cancer cell clusters without the formation of any gland-like structure, whereas intestinal GC is featured by cohesive cells that form glandular structure, whose histology and morphology are similar to intestinal adenocarcinoma [6]. There is a broad consensus that diffuse GC and intestinal GC are two distinct diseases with different molecular base, etiology, and epidemiology, which may also benefit from different therapeutic approaches [7]. For diffuse GC, which is equally distributed between males and females, the incidence rates are similar in all geographic locations [1,8]. Loss of expression of E-cadherin, by mutation or hypermethylation, occurred in nearly 90% of diffuse GCs [3,9,10]. Intestinal GC, which is more common in males, is highly prevalent especially in Eastern Asia [1,8]. *Helicobacter pylori* infection is the most important risk factor of intestinal GC, which resulted in a sequence of molecular events (atrophic gastritis, intestinal metaplasia, dysplasia, intestinal GC) [8,11,12]. In addition, diffuse GC is linked to familial occurrence and got a more unfavorable prognosis compared with intestinal GC.

It is also illustrated that some genes acted differentially in diffuse GC and intestinal GC [7,13,14]; however, only a few prognostic biomarkers for specific subtype GC have been discovered. Several pilot studies showed transcriptome level difference between these two subtypes [7,10,15–17], yet large-scale, systematic, and comprehensive investigation of gene expression difference between diffuse GC and intestinal GC based on large populations is still needed to shed light on precious medication on different GC patients.

In this study, microarray data of a large cohort of GC patients with long-time follow-up were collected, and integrated analysis of several bioinformatics tools was applied to reveal the molecular profile of these two GC subtypes and seek for subtype-specific prognostic biomarkers.

Materials and Methods

Data Sources and Preprocessing

cDNA microarray datasets GSE62254 and GSE15459 were downloaded from the GEO Web site. All sample information with Lauren classification and long-time follow-up included was downloaded from the original articles [18,19]. Robust multichip average method [20] was used for background correction, and qspline method was for normalization [21]. Datasets were then PM (Perfect Match)-corrected by using only perfect match and summarized by the Li-Wong model [22]. All probes were mapped to Ensembl Gene Symbols by R package mygene [23].

Identification of Differentially Expressed Genes between Subgroups

Patients were divided into three groups (diffuse GC, mixed GC, and intestinal GC) according to the Lauren classification. Kolmogorov-Smirnov test was applied to test whether data are normally distributed in each subgroup. Diffuse specifically expressed genes were determined according to three conditions: 1) for mean expression, diffuse GC > mixed GC > intestinal GC; 2) diffuse GC versus intestinal GC Student's *t* test $P < .05$ and false discovery rate (FDR) < 0.25; 3) for mean expression, diffuse GC > 2× intestinal GC. Similar standards were also set up for the definition of intestinal specific expressed genes (see Figure 1 and “Results” section for details).

Gene Ontology (GO) Enrichment Analysis

GO enrichment analysis of biological process category was performed by R package “GStats.” Hypergeometric distribution model along with FDR adjustment was used for significance evaluation. All significantly

enriched GOs had been reviewed, and the representative GOs for both groups were selected by removing par synonymous and redundant terms.

Survival Analysis

Overall survival, defined as the time from surgery to death/last follow-up, served as the primary end point. For each gene, the median expression level was chosen as the cutoff to divide subgroups. Log-rank test was applied to reveal the difference between survival of two subgroups, and a Kaplan-Meier plot was drawn to show the results intuitively. Cox proportional hazard model was implemented for multiple-variants analysis. Age, sex, and tumor-node-metastasis stage were all included in the model at the beginning step, and backward LR stepwise logistic regression was performed for variable selection.

Statistical Analysis and R Package Usage

All analyses were performed by R Software 3.2.2 (www.r-project.org). Expression of each gene between two groups was compared by Student's *t* test, and FDR was utilized to correct for multiple testing. *P* value < .05 was considered statistically significant. R package qvalue was used for FDR analysis, and a cutoff of FDR < 0.25 was selected.

Results

Difference between Diffuse and Intestinal GC at Same Stage, Early and Advanced GC with Same Lauren Classification

All 300 patients from the derivation cohort (GSE62254) were stratified into two subgroups according to their clinical stage (stage I/II and stage III/IV). There were 4053 genes differentially expressed between diffuse and intestinal GC in the early stage, whereas 6161 genes were differentially expressed between diffuse and intestinal GC in the advanced stage. When stratified by Lauren classification, there were only 27 genes differentially expressed between early and advanced stages in diffuse GC and 17 genes between early and advanced stages in intestinal GC (Figure 2, A). Differentially expressed genes between diffuse and intestinal GCs in the same stage were much more than those between early and advanced GCs with the same Lauren classification, which strongly indicated that diffuse and intestinal GCs were two distinct cancer subtype not only in morphology level but also in molecular level.

Highly Differentially Expressed Genes between Diffuse and Intestinal GC

When we merged patients of early and advanced stage together, 4598 genes were found differentially expressed between diffuse and intestinal GC: 1752 were specifically highly expressed in diffuse GC (mean expression: diffuse GC > mixed GC > intestinal GC), and 2846 were specifically highly expressed in intestinal GC (mean expression: intestinal GC > mixed GC > diffuse GC). GO biological process ontology was enriched for both diffuse and intestinal GC patients. Immune and inflammatory response, cell adhesion, Rho GTPase, and angiogenesis-related ontologies were enriched for diffuse GC, whereas DNA repair, cell cycle, and p53-related ontologies were enriched for intestinal GC. Apoptosis and metabolic process were enriched for both diffuse and intestinal GCs. Representative GOs enriched for both diffuse and intestinal subgroups, along with their odds ratios (ORs) and *P* values, were displayed in Figure 2, C and D and Table 1.

To address the most differentially expressed genes between diffuse and intestinal GC, a two-fold threshold of expression change (along with $P < .05$ and FDR < 0.25) was used as a comprehensive standard to

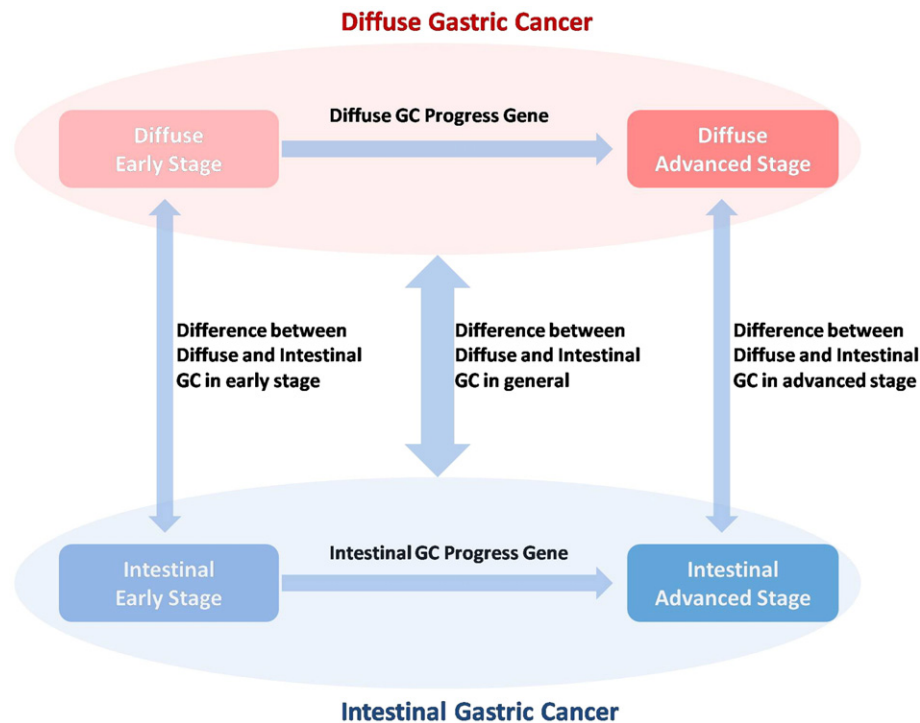


Figure 1. Conceptual summary of all subgroups and their mutual comparisons.

filter highly differentially expressed genes between diffuse and intestinal GCs. Thirty-six genes were found specifically highly expressed in diffuse GC, and 4 genes were found specifically highly expressed in intestinal GC. Thirty-nine of these 40 specifically expressed genes were also differentially expressed between diffuse and intestinal GCs at the advanced stage (Figure 2, B).

Diffuse and Intestinal Specifically Expressed Genes' Signature for GC Patients

Using the 40 specifically expressed genes as a signature, 300 GC patients could be divided into three subgroups by hierarchical clustering based on its bifurcations (Figure 3, A; DF: diffuse, MX: mixed, IT: intestinal). Cluster 1 patients (DF:MX:IT = 51:15:106) with intestinal-like molecular signature had the best prognosis, whereas cluster3 patients (DF:MX:IT = 31:1:5) with diffuse-like molecular signature had the worst prognosis. Log-rank test indicated that the difference of survival among three clusters was statistically significant ($\chi^2 = 30.71$, $P < .001$). Besides, clustering by the given signature showed better discrimination in prognosis than grouping by Lauren classification (Figure 3, B and C).

To validate the prognostic power of the 40-gene signature, GSE15459 was recruited as a validation cohort. When stratified by the 40-gene signature, cluster 1 patients (DF:MX:IT = 3:2:16) with intestinal-like molecular signature had the best prognosis, whereas cluster 3 patients (DF:MX:IT = 58:8:41) with diffuse-like molecular signature had the worst prognosis ($\chi^2 = 4.81$, $P = .090$; Figure 3, D and E). However, there was no significant difference among diffuse, mixed, and intestinal GCs in this cohort ($\chi^2 = 1.09$, $P = .557$, Figure 3, F).

For both cohorts, the given 40-gene signature had better prognostic discrimination than the Lauren classification.

Identification of Diffuse and Intestinal Specific Prognostic Biomarkers

Among the 40 genes specifically expressed in diffuse or intestinal GC, HSPB8, SYNPO2, ACTG2, and SCRG1 were found to be

prognostic factors in both diffuse and intestinal GCs (data not shown). MGP, THBS4, EFEMP1, FRZB, GGTA1P, and C2orf40 were identified as prognostic factors specifically in diffuse GC but not in intestinal GC (Figure 4, A, C, and E; Supplementary Figure 1, A, B, and C), whereas KRT23 was identified as a prognostic factor specifically in intestinal GC but not in diffuse GC (Figure 4, G and H). Prognostic value of all the above genes in diffuse and intestinal GC was also investigated in the validation cohort, and similar results were obtained (Figure 4, B, D, and F; Supplementary Figure 1, D, E, and F). Combined with clinical pathological data, Cox proportional hazards models were respectively implemented for MGP, THBS4, EFEMP1, FRZB, GGTA1P, C2orf40, and KRT23. MGP [RR (95% CI) = 1.644 (1.018-2.657), $P = .042$], EFEMP1 [RR (95% CI) = 1.537 (0.969-2.437), $P = .067$] and FRZB [RR (95% CI) = 1.824 (1.115-2.986), $P = .017$] were identified as independent prognostic factors in diffuse GC, whereas KRT23 [RR (95% CI) = 1.616(0.938-2.785), $P = .083$] was identified as an independent prognosis factor in intestinal GC (results of EFEMP1 and KRT23 shown in Tables 2 and 3). The independent prognostic roles of EFEMP1 [RR (95% CI) = 2.464 (1.248-4.865), $P = .009$], FRZB [RR (95% CI) = 2.157 (1.143-4.069), $P = .017$], and KRT23 [RR (95% CI) = 3.131(1.674-5.854), $P < .001$] were also verified in validation cohort. However, MGP failed to be identified as an independent prognostic factor of diffuse GC in the validation cohort (results of EFEMP1 and KRT23 shown in Supplementary Tables 1 and 2).

Discussion

Molecular signatures associated with distinct prognosis have been illustrated in many different cancers [24–28]. For breast cancer, a 21-gene signature has even been approved by the FDA for application in bedside decisions [24,29]. Various studies tried to discover new classifier for GC, but none of them had enough potential to be used in clinical practice. For example, Cristescu et al. merged both CNV chip and cDNA chip data and proposed that EMT, MSI, and TP53

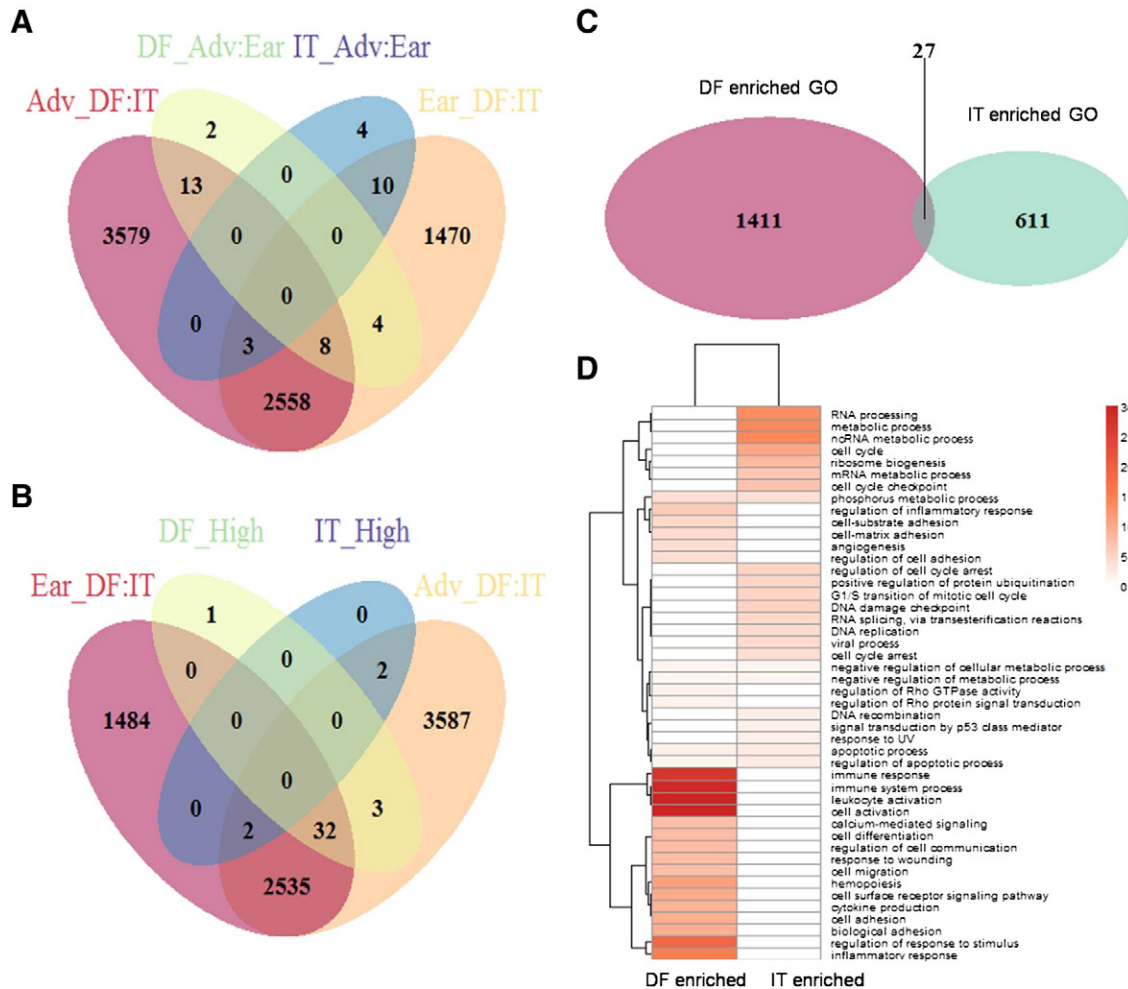


Figure 2. Specifically expressed genes of diffuse and intestinal GC and their enriched GO analysis. (A) Venn diagram of four kinds of differentially expressed genes (Adv_DF:IT, advanced diffuse vs. advanced intestinal; Ear_DF:IT, early diffuse vs. early intestinal; DF_Adv:Ear, advanced diffuse vs. early diffuse; IT_Adv:Ear, advanced intestinal vs. early intestinal). (B) Venn diagram of four kinds of differentially expressed genes (DF_High, diffuse GC-specific highly expressed genes; IT_High, intestinal GC-specific highly expressed genes). Venn diagram (C) and heatmap (D) of DF enriched GOs and IT enriched GOs.

activity could be integrated to define GC molecular subtypes [18], which had a considerable discrimination of prognosis. However, such a classification method that integrated both CNV chip and cDNA chip data was much less cost-effective than regular gene signatures. Tanabe et al. suggested that using only two genes could distinguish diffuse GC from intestinal GC and identify the poor-prognosis subgroup [16], yet this method showed no advantage over the Lauren system and the inner heterogeneity in intestinal GC/diffuse GC was ignored.

Based on the macroscopic appearance of GC, the Lauren classification system has been popular in clinicians for decades due to its robustness and ease of implementation [4]. It is commonly accepted that diffuse GC and intestinal GC are two distinct diseases in all aspects and that intestinal GC has a better prognosis compared with diffuse GC [3,18]. However, actually, the Lauren classification is not a good subtyping method for prognosis. In this study, we revealed that the difference among diffuse, mixed, and intestinal GCs is statistically significant in GSE62254 but not in GSE15459. Possible reasons for this distinction would be the following: 1) The

discrimination of diffuse, mixed, and intestinal GCs was mainly based on pathological features which would be affected by the subjective judgment of pathologists. 2) The ethnicity of GSE62254 (a mixed population that involved Japanese, Korean, Chinese, and American Asians) is different with that of GSE15459 (Singapore Asians) [18,19].

Differences on transcriptomic level between diffuse and intestinal GC had been investigated for many studies [7,10,15–17], but all of these studies were based on a relatively small population. Gene lists generated by these previous studies were merely overlapping. Additionally, no survival data were incorporated into these studies, which made their conclusions less confident. Thus, a large sample paired with clinical information was needed to evaluate the clinical-translational diversity between the two GC subtypes.

In this study, differentially expressed genes in diffuse GC and intestinal GC were identified based on a population of 300 GC patients [18]. Cell adhesion- and Rho GTPase-related genes were enriched in diffuse GC, which could be interpreted by the high mutation and methylation level of CDH1 [10] and the high mutation

Table 1. Representative GOs Enriched in Diffuse and Intestinal GC

GOBPID	DF Enriched GOs		IT Enriched GOs		Term
	P Value	OR	P Value	OR	
GO: 0045321	1.27E-33	3.707469	0.999973	0.592131	Leukocyte activation
GO: 0001775	3.05E-31	3.121759	0.999999	0.586179	Cell activation
GO: 0002376	7.09E-30	2.240386	0.999995	0.737616	Immune system process
GO: 0006955	1.44E-27	2.504605	1	0.59028	Immune response
GO: 0048583	1.30E-19	1.81444	0.99987	0.805013	Regulation of response to stimulus
GO: 0006954	6.07E-16	2.674378	0.999251	0.653624	Inflammatory response
GO: 0030097	7.02E-12	2.229078	0.999248	0.677301	Hemopoiesis
GO: 0007166	1.28E-10	1.524782	1	0.64234	Cell surface receptor signaling pathway
GO: 0007155	3.07E-10	1.860816	0.99998	0.664989	Cell adhesion
GO: 0022610	4.31E-10	1.849946	0.999976	0.668609	Biological adhesion
GO: 0001816	1.03E-09	2.196095	0.998834	0.660077	Cytokine production
GO: 0030154	9.59E-09	1.457505	0.999995	0.773168	Cell differentiation
GO: 0010646	9.79E-09	1.497657	0.996572	0.844485	Regulation of cell communication
GO: 0009611	1.46E-08	1.802479	0.995692	0.767135	Response to wounding
GO: 0016477	1.98E-08	1.763924	0.987177	0.806766	Cell migration
GO: 0019722	2.84E-08	3.871589	0.715387	0.877997	Calcium-mediated signaling
GO: 0050727	8.88E-07	2.458491	0.997962	0.547671	Regulation of inflammatory response
GO: 0031589	2.42E-05	2.113268	0.806245	0.865792	Cell-substrate adhesion
GO: 0007160	4.82E-05	2.358447	0.902885	0.757264	Cell-matrix adhesion
GO: 0001525	6.31E-05	1.857824	0.999253	0.59139	Angiogenesis
GO: 0030155	6.80E-05	1.928391	0.944933	0.770372	Regulation of cell adhesion
GO: 0032319	0.018652	1.683671	0.789575	0.848437	Regulation of Rho GTPase activity
GO: 0035023	0.024772	1.595994	0.79659	0.851779	Regulation of Rho protein signal transduction
GO: 0071156	0.36453	1.173835	6.44E-06	2.797936	Regulation of cell cycle arrest
GO: 0016032	0.365411	1.055653	5.92E-05	1.488715	Viral process
GO: 0007050	0.549637	0.991216	0.000104	1.883741	Cell cycle arrest
GO: 0000082	0.915469	0.720671	8.47E-06	2.020819	G1/S transition of mitotic cell cycle
GO: 0072331	0.934457	0.639055	0.009177	1.643644	Signal transduction by p53 class mediator
GO: 0031398	0.966859	0.546816	1.50E-05	2.345111	Positive regulation of protein ubiquitination
GO: 0009411	0.971164	0.489318	0.009073	1.786106	Response to UV
GO: 0006310	0.974806	0.599477	0.006565	1.557605	DNA recombination
GO: 0000375	0.987431	0.546555	3.01E-05	1.967992	RNA splicing
GO: 0006260	0.987624	0.596876	5.47E-05	1.781074	DNA replication
GO: 0008152	0.991442	0.866837	8.78E-15	1.486394	Metabolic process
GO: 0000077	0.996557	0.371127	6.16E-06	2.396237	DNA damage checkpoint
GO: 0016071	0.998252	0.624284	1.57E-07	1.71282	mRNA metabolic process
GO: 0042254	0.999066	0.325967	7.64E-09	2.805316	Ribosome biogenesis
GO: 0006396	0.999873	0.559105	5.73E-14	2.053407	RNA processing
GO: 0000075	0.999916	0.314511	8.47E-08	2.280505	Cell cycle checkpoint
GO: 0007049	0.999998	0.613873	4.27E-11	1.586689	Cell cycle
GO: 0034660	0.999999	0.264909	1.44E-14	2.728084	ncRNA metabolic process

rate of RhoA gene [30,31]. Cell cycle and p53-DNA repair-related genes were found enriched in intestinal GC, which is in coincidence with previous reports [15,32]. By stringently defining the term of “specific expressed genes” in both subgroups, a 40-gene signature was generated. In both derivation cohort and validation cohort, the 40-gene signature showed better discrimination for prognosis than the Lauren classification, indicating that these genes revealed the deep-seated molecular diversities between different GC patients.

By evaluating the 40 genes in the signature, 3 genes were identified as GC subtype-specific biomarkers. EFEMP1 and FRZB were both independent prognostic factors in diffuse GC but not intestinal GC, whereas KRT23 was an independent prognostic factor in intestinal GC but not diffuse GC. EFEMP1 encodes an extracellular matrix glycoprotein, which was upregulated in glioma and may play a role in its aggressiveness [33,34]. FRZB encodes a secreted protein that is involved in the regulation of bone development, which could also influence Wnt/ β -catenin signaling in GC [35,36]. Our findings indicated that EFEMP1 and FRZB may also be involved in diffuse GC-specific pathways, such as cell adhesion. KRT23 encodes a member of the keratins, which is responsible for the structural

integrity of epithelial cells. It is not surprising that KRT23 played an important role in intestinal GC considering that it was also found as an oncogene which could affect proliferation and DNA damage response of colon cancer cells [37,38].

In conclusion, based on large GC patient cohorts, we found that GCs with different Lauren classifications had different molecular characteristics, and then we interpreted both profiles in search of insights into the mechanisms underlying the biological behavior. A 40-gene signature was proposed, which had a better discrimination than the Lauren system for prognosis. For both diffuse and intestinal GCs, we identified several specific biomarkers (EFEMP1 and FRZB for diffuse GC; KRT23 for intestinal GC) which not only displayed very promising prognostic potential but also shed light on the further exploration of the deep difference between GC subtypes.

Supplementary data to this article can be found online at [doi:10.1016/j.tranon.2016.11.003](https://doi.org/10.1016/j.tranon.2016.11.003).

Competing Interests

The authors declare that they have no competing interests.

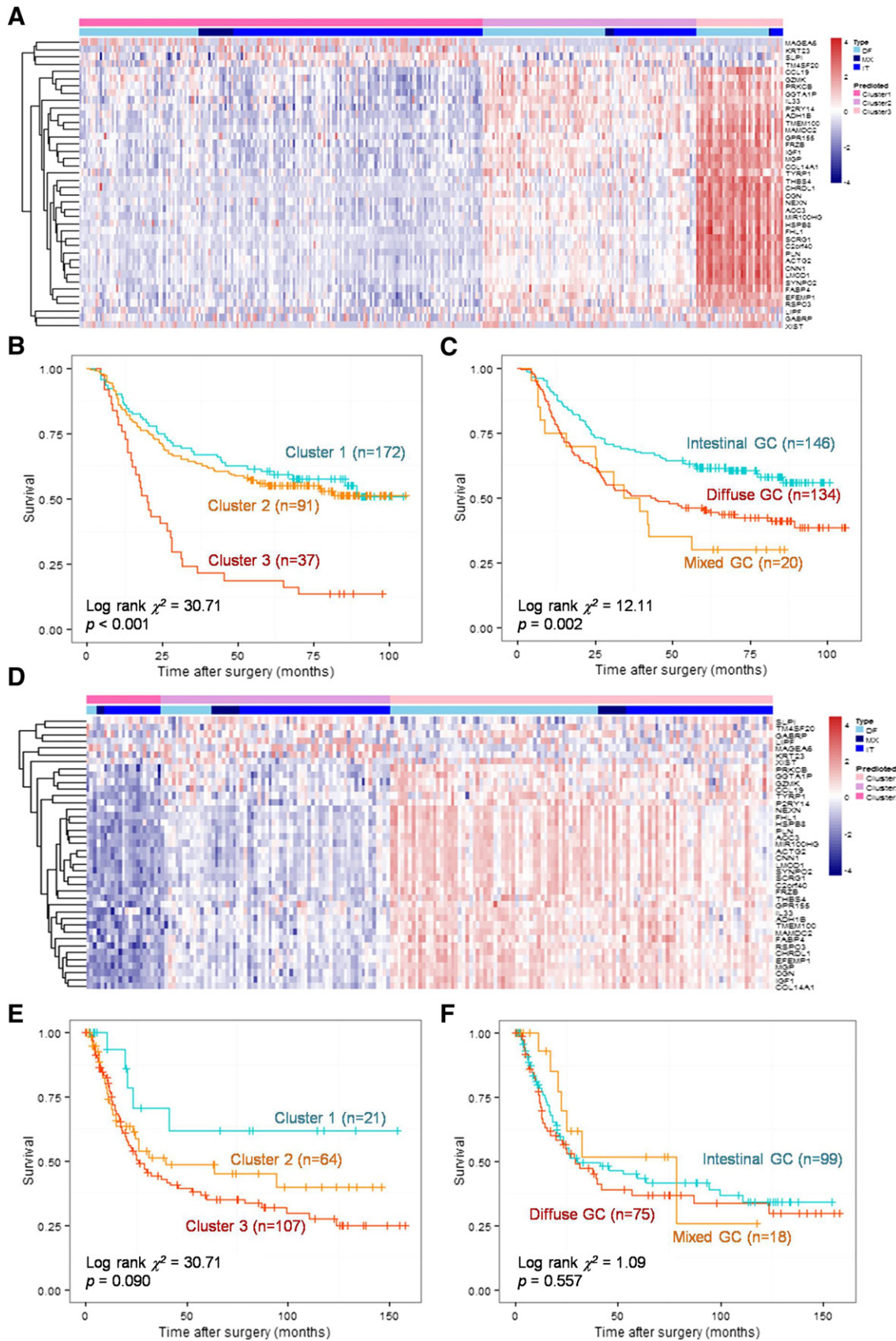


Figure 3. Kaplan-Meier plots for diffuse- and intestinal-specific prognostic biomarkers. Hierarchical clustering results of 40-gene signature (A) in the derivation cohort. A Kaplan-Meier curve for grouping by 40-gene signature (B) and Lauren classification (C) in the derivation cohort. Hierarchical clustering results of 40-gene signature (D) in the validation cohort. A Kaplan-Meier curve for grouping by 40-gene signature (E) and Lauren classification (F) in the validation cohort.

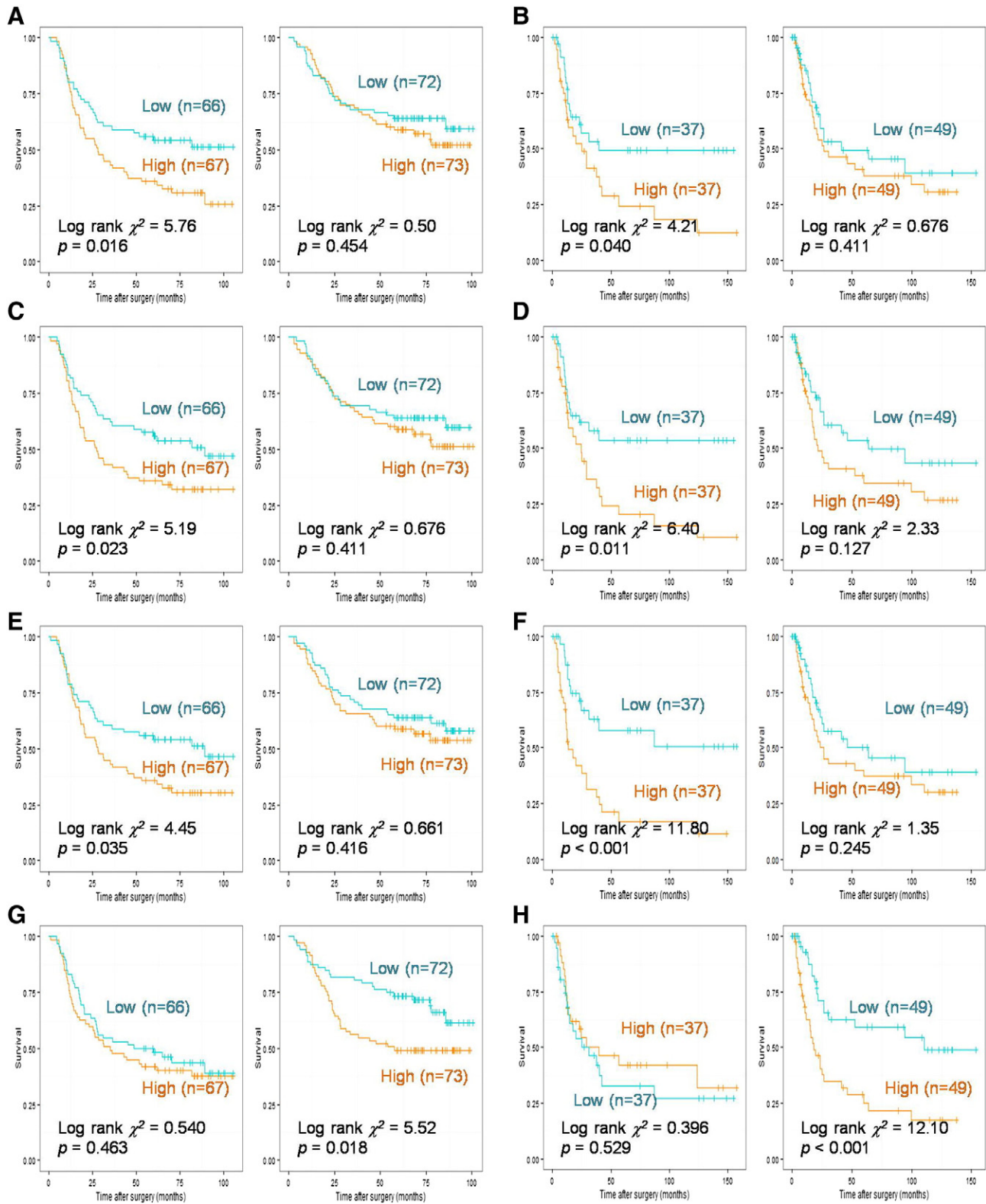


Figure 4. Kaplan-Meier curves for diffuse- and intestinal-specific prognostic biomarkers. A Kaplan-Meier curve for gene MGP (AB), THBS4 (CD), and EFEMP1 (EF), KRT23 (GH) in the derivation cohort (ACEG) and the validation cohort (BDFH). For each gene, a plot for diffuse GC is displayed on the left panel and a plot for intestinal GC on the right panel. A cutoff was made on the median value of each subgroup.

Authors' Contributions

L. M. and Y. Z. carried out the main analysis. L. M. and S. G. conceived and designed the study. R. C. and J. X. helped to collect

and reformat the primary data. X. Q. and S. Z. helped to analyze data and revise the manuscript. L. M., L. S., and S. Z. drafted the manuscript. All authors read and approved the final manuscript.

Table 2. Independent Prognostic Value of EFEMP1 in Diffuse and Intestinal GC

		Diffuse GC		Intestinal GC	
		HR (95% CI)	P Value	HR (95% CI)	P Value
EFEMP1	Pos vs neg	1.537 (0.969-2.437)	.067	Variable eliminated	
Sex	M vs F	Variable eliminated		2.089 (1.042-4.186)	.038
Age		1.023 (1.003-1.042)	.021	1.058 (1.022-1.095)	.001
Stage	Stage III vs I/II	1.555 (0.689-3.507)	.287	Variable eliminated	
	Stage IV vs I/II	7.440 (2.756-20.085)	<.001	Variable eliminated	
T	III, IV vs I, II	Variable eliminated		3.285 (1.906-5.665)	<.001
N	Pos vs neg	1.314 (0.899-1.918)	.158	1.510 (1.111-2.054)	.008
M	Pos vs neg	Variable eliminated		2.448 (0.826-7.260)	.106
Overall Cox model			<.001		<.001

Table 3. Independent Prognostic Value of KRT23 in Diffuse and Intestinal GC

		Diffuse GC		Intestinal GC	
		HR (95% CI)	P Value	HR (95% CI)	P Value
KRT23	Pos vs neg	Variable eliminated		1.616 (0.938-2.785)	.083
Sex	M vs F	Variable eliminated		2.091 (1.033-4.234)	.04
Age		1.022 (1.003-1.042)	.022	1.059 (1.023-1.096)	.001
Stage	Stage III vs I/II	2.131 (1.012-4.488)	.046	Variable eliminated	
	Stage IV vs I/II	12.619 (6.037-26.377)	<.001	Variable eliminated	
T	III, IV vs I, II	Variable eliminated		3.062 (1.762-5.323)	<.001
N	Pos vs neg	Variable eliminated		1.442 (1.060-1.961)	.019
M	Pos vs neg	Variable eliminated		2.899 (0.959-8.764)	.059
Overall Cox model			<.001		<.001

Acknowledgement

We thank Prof. Chengchao Shou and Prof. Like Qu (Peking University) for their valuable advices to this study. We thank Dr. Finci (Tsinghua University) for the meticulous editing of this manuscript. We thank the Clinical Data and Biobank Resource of Beijing Friendship Hospital for their help in data processing.

References

- Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, and Jemal A (2015). Global cancer statistics, 2012. *CA Cancer J Clin* **65**, 87–108.
- Siegel RL, Miller KD, and Jemal A (2015). Cancer statistics, 2015. *CA Cancer J Clin* **65**, 5–29.
- Tan P and Yeoh K-G (2015). Genetics and molecular pathogenesis of gastric adenocarcinoma. *Gastroenterology* **149**, 1153–1162 [e1153].
- Lauren T (1965). The two histological main types of gastric carcinoma, an attempt at a histoclinical classification. *Acta Pathol Microbiol Scand* **64**, 19.
- Hartgrink HH, Jansen JPM, Grieken N, and van de Velde CJH (2009). Gastric cancer. *Lancet* **374**, 14.
- Shah MA, Khanin R, Tang L, Janjigian YY, Klimstra DS, Gerdes H, and Kelsen DP (2011). Molecular classification of gastric cancer: a new paradigm. *Clin Cancer Res* **17**, 2693–2701.
- Förster S, Gretschel S, Jöns T, Yashiro M, and Kemmner W (2011). THBS4, a novel stromal molecule of diffuse-type gastric adenocarcinomas, identified by transcriptome-wide expression profiling. *Mod Pathol* **24**, 1390–1403.
- Fock KM (2014). Review article: the epidemiology and prevention of gastric cancer. *Aliment Pharmacol Ther* **40**, 250–260.
- Oh J-H, Jung S-H, Hong S-J, and Rhyu M-G (2015). DNA methylation as surrogate marker for gastric cancer. *J Cancer Prev* **20**, 7.
- Lee YS, Cho YS, Lee GK, Lee S, Kim KW, Jho S, Kim HM, and Hong SH (2014). Genomic profile analysis of diffuse-type gastric cancers. *Genome Biol* **15**, 15.
- Kupcinkas J, Wex T, Link A, Leja M, Bruzaite I, Steponaitiene R, Juzenas S, Gyvyte U, Ivanauskas A, and Ancans G, et al (2014). Gene polymorphisms of microRNAs in *Helicobacter pylori*-induced high risk atrophic gastritis and gastric cancer. *PLoS One* **9**, e87467.
- Chaturvedi R, de Sablet T, Asim M, Piazuolo MB, Barry DP, Verriere TG, Sierra JC, Hardbower DM, Delgado AG, and Schneider BG, et al (2015). Increased *Helicobacter pylori*-associated gastric cancer risk in the Andean region of Colombia is mediated by spermine oxidase. *Oncogene* **34**, 3429–3440.
- Marrelli D, Polom K, Pascale V, Vindigni C, Piagnerelli R, De Franco L, Ferrara F, Roviello G, Garosi L, and Petrioli R, et al (2016). Strong prognostic value of microsatellite instability in intestinal type non-cardia gastric cancer. *Ann Surg Oncol* **23**, 943–950.
- Franke K, Carl-McGrath S, Rohl FW, Lendeckel U, Ebert MP, Tanzer M, Pross M, and Rocken C (2009). Differential expression of SPARC in intestinal-type gastric cancer correlates with tumor progression and nodal spread. *Transl Oncol* **2**, 310–320.
- Jinawath N, Furukawa Y, Hasegawa S, and Li M (2004). Comparison of gene-expression profiles between diffuse- and intestinal-type gastric cancers using a genome-wide cDNA microarray. *Oncogenes* **23**, 15.
- Tanabe S, Aoyagi K, Yokozaki H, and Sasaki H (2014). Gene expression signatures for identifying diffuse-type gastric cancer associated with epithelial-mesenchymal transition. *Int J Oncol* **44**, 16.
- Kim B, Bang S, Lee S, Kim S, and Jung Y (2003). Expression profiling and subtype-specific expression of stomach cancer. *Cancer Res* **63**, 9.
- Cristescu R, Lee J, Nebozhyn M, Kim K-M, Ting JC, Wong SS, Liu J, Yue YG, Wang J, and Yu K, et al (2015). Molecular analysis of gastric cancer identifies subtypes associated with distinct clinical outcomes. *Nat Med* **21**, 449–456.
- Chia N-Y, Deng N, Das K, Huang D, Hu L, Zhu Y, Lim KH, Lee M-H, Wu J, and Sam XX, et al (2015). Regulatory crosstalk between lineage-survival oncogenes KLF5, GATA4 and GATA6 cooperatively promotes gastric cancer development. *Gut* **64**, 707–719.
- Irizarry RA, Hobbs B, Collin F, Beazer YD, and Antonellis KJ (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Bioinformatics* **4**, 249–264.
- Workman C, Jensen LJ, Jarmer H, Berka R, and Gautier L (2002). A new non-linear normalization method for reducing variability in DNA microarray experiments. *Genome Biol* **3** [0048.0041–0048.0016].
- Li C (2000). Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. *Proc Natl Acad Sci* **98**, 31–36.
- Wu C, MacLeod I, and Su AI (2012). BioGPS and MyGene.info: organizing online, gene-centric information. *Nucleic Acids Res* **41**, D561–D565.
- Nielsen T, Wallden B, Schaper C, Ferree S, Liu S, Gao D, Barry G, Dowidar N, Maysuria M, and Storhoff J (2014). Analytical validation of the PAM50-based Prosigna Breast Cancer Prognostic Gene Signature Assay and nCounter Analysis

- System using formalin-fixed paraffin-embedded breast tumor specimens. *BMC Cancer* **14**, 177.
- [25] Katayama H, Boldt C, Ladd JJ, Johnson MM, Chao T, Capello M, Suo J, Mao J, Manson JE, and Prentice R, et al (2015). An autoimmune response signature associated with the development of triple negative breast cancer reflects disease pathogenesis. *Cancer Res*.
- [26] Binato R, de Almeida Oliveira NC, Du Rocher B, and Abdelhay E (2015). The molecular signature of AML mesenchymal stromal cells reveals candidate genes related to the leukemogenic process. *Cancer Lett* **369**, 134–143.
- [27] Merlos-Suárez A, Barriga Francisco M, Jung P, Iglesias M, Céspedes María V, Rossell D, Sevillano M, Hernando-Momblona X, da Silva-Diz V, and Muñoz P, et al (2011). The intestinal stem cell signature identifies colorectal cancer stem cells and predicts disease relapse. *Cell Stem Cell* **8**, 511–524.
- [28] Wang L, Shen X, Wang Z, Xiao X, Wei P, Wang Q, Ren F, Wang Y, Liu Z, and Sheng W, et al (2015). A molecular signature for the prediction of recurrence in colorectal cancer. *Mol Cancer* **14**, 22.
- [29] Alvarado MD, Prasad C, Rothney M, Cherbavaz DB, Sing AP, Baehner FL, Svedman C, and Markopoulos CJ (2015). A prospective comparison of the 21-gene recurrence score and the PAM50-based Prosigna in estrogen receptor-positive early-stage breast cancer. *Adv Ther*.
- [30] Yoon C, Cho SJ, Aksoy BA, Park DJ, Schultz N, Ryeom S, and Yoon SS (2015). Chemotherapy resistance in diffuse type gastric adenocarcinoma is mediated by RhoA activation in cancer stem-like cells. *Clin Cancer Res*.
- [31] Kakiuchi M, Nishizawa T, Ueda H, Gotoh K, Tanaka A, Hayashi A, Yamamoto S, Tatsuno K, Katoh H, and Watanabe Y, et al (2014). Recurrent gain-of-function mutations of RHOA in diffuse-type gastric carcinoma. *Nat Genet* **46**, 583–587.
- [32] Zheng H, Takahashi H, Murai Y, Cui Z, Nomoto K, Miwa S, Tsuneyama K, and Takano Y (2006). Pathobiological characteristics of intestinal and diffuse-type gastric carcinoma in Japan: an immunostaining study on the tissue microarray. *J Clin Pathol* **60**, 273–277.
- [33] Zhang S, Ye Z, Song X, Chen G, Huai C, Wang Q, Song J, Lu D, Zhao Y, and Chen H (2015). Association of EFEMP1 gene polymorphisms with the risk of glioma: a hospital-based case-control study in a Chinese Han population. *J Neurol Sci* **349**, 54–59.
- [34] Almeida M, Costa VL, Costa NR, Ramalho-Carvalho J, Baptista T, Ribeiro FR, Paulo P, Teixeira MR, Oliveira J, and Lothe RA, et al (2014). Epigenetic regulation of EFEMP1 in prostate cancer: biological relevance and clinical potential. *J Cell Mol Med* **18**, 2287–2297.
- [35] Schiefer L, Visweswaran M, Perumal V, Arfuso F, Groth D, Newsholme P, Warriar S, and Dharmarajan A (2014). Epigenetic regulation of the secreted frizzled-related protein family in human glioblastoma multiforme. *Cancer Gene Ther* **21**, 297–303.
- [36] Qin S, Zhang Z, Li J, and Zang L (2014). FRZB knockdown upregulates beta-catenin activity and enhances cell aggressiveness in gastric cancer. *Oncol Rep* **31**, 2351–2357.
- [37] Birkenkamp-Demtroder K, Hahn SA, Mansilla F, Thorsen K, Maghnoji A, Christensen R, Oster B, and Orntoft TF (2013). Keratin23 (KRT23) knockdown decreases proliferation and affects the DNA damage response of colon cancer cells. *PLoS One* **8**, e73593.
- [38] Birkenkamp-Demtroder K, Mansilla F, Sorensen FB, Kruhoffer M, Cabezon T, Christensen LL, Aaltonen LA, Verspaget HW, and Orntoft TF (2007). Phosphoprotein keratin 23 accumulates in MSS but not MSI colon cancers in vivo and impacts viability and proliferation in vitro. *Mol Oncol* **1**, 181–195.