




Epigenome-Wide Analysis of Methylation Changes in the Sequence of Gallstone Disease, Dysplasia, and Gallbladder Cancer

Johannes Brägelmann ,^{1,3} Carol Barahona Ponce,^{1,4} Katherine Marcelain,⁴ Stephanie Roessler,⁵ Benjamin Goeppert,⁵ Ivan Gallegos,⁶ Alicia Colombo,^{4,6} Verónica Sanhueza,⁷ Erik Morales,⁸ María Teresa Rivera,⁹ Gonzalo de Toro,¹⁰ Alejandro Ortega,¹¹ Bettina Müller,¹² Fernando Gabler,¹³ Dominique Scherer,¹ Melanie Waldenberger,¹⁴ Eva Reischl,¹⁴ Felix Boekstegers ,¹ Valentina Garate-Calderon,^{1,4} Sinan U. Umu,¹⁵ Trine B. Rounge,^{15,16} Odilia Popanda,¹⁷ and Justo Lorenzo Bermejo ¹

BACKGROUND AND AIMS: Gallbladder cancer (GBC) is a highly aggressive malignancy of the biliary tract. Most cases of GBC are diagnosed in low-income and middle-income countries, and research into this disease has long been limited. In this study we therefore investigate the epigenetic changes along the model of GBC carcinogenesis represented by the sequence gallstone disease → dysplasia → GBC in Chile, the country with the highest incidence of GBC worldwide.

APPROACH AND RESULTS: To perform epigenome-wide methylation profiling, genomic DNA extracted from sections of formalin-fixed, paraffin-embedded gallbladder tissue was analyzed using Illumina Infinium MethylationEPIC BeadChips. Preprocessed, quality-controlled data from 82 samples (gallstones $n = 32$, low-grade dysplasia $n = 13$, high-grade dysplasia $n = 9$, GBC $n = 28$) were available to identify differentially methylated markers, regions, and pathways as well as changes in copy number variations (CNVs). The number and magnitude of epigenetic changes increased with disease development and predominantly involved the hypermethylation of cytosine-guanine dinucleotide islands and gene promoter regions. The methylation of genes implicated

in Wnt signaling, Hedgehog signaling, and tumor suppression increased with tumor grade. CNVs also increased with GBC development and affected cyclin-dependent kinase inhibitor 2A, MDM2 proto-oncogene, tumor protein P53, and cyclin D1 genes. Gains in the targetable Erb-B2 receptor tyrosine kinase 2 gene were detected in 14% of GBC samples.

CONCLUSIONS: Our results indicate that GBC carcinogenesis comprises three main methylation stages: early (gallstone disease and low-grade dysplasia), intermediate (high-grade dysplasia), and late (GBC). The identified gradual changes in methylation and CNVs may help to enhance our understanding of the mechanisms underlying this aggressive disease and eventually lead to improved treatment and early diagnosis of GBC. (HEPATOLOGY 2021;73:2293-2310).

Gallbladder cancer (GBC; *International Classification of Diseases*, Tenth Revision, diagnosis code C23) is a highly aggressive malignancy that accounts for 80%–95% of biliary tract cancers and every year affects more than 175,000 persons

Abbreviations: ANK1, ankyrin 1; ANOVA, analysis of variance; APC, APC regulator of WNT signaling pathway; 5-aza-dC, 5'-aza-2'-deoxycytidine; CCA, cholangiocarcinoma; CCND1, cyclin D1; CDCA7, cell division cycle associated 7; CDK, cyclin-dependent kinase; CDKN2A, cyclin-dependent kinase inhibitor 2A; CDO1, cysteine dioxygenase type 1; CNV, copy number variation; CpG, cytosine-guanine dinucleotide; de-CISH, dual-color chromogenic in situ hybridization; DMR, differentially methylated region; DMSO, dimethyl sulfoxide; EDNRB, endothelin receptor type B; ERBB2, Erb-B2 receptor tyrosine kinase 2; FEZF2, FEZ family zinc finger 2; FFPE, formalin-fixed, paraffin-embedded; FGF, fibroblast growth factor; FGFR3, FGF receptor 3; GBC, gallbladder cancer; GSD, gallstone disease; HAND2, heart and neural crest derivatives expressed 2; HBE1, hemoglobin subunit epsilon 1; HHIP, hedgehog interacting protein; HMGAI, high-mobility group AT-hook 1; IHC, immunohistochemistry; MDM2, MDM2 proto-oncogene; MET, MET proto-oncogene, receptor tyrosine kinase; PCA, principal component analysis; RNAseq, RNA sequencing; RUNX3, RUNX family transcription factor 3; TP73, tumor protein P73; TRIM57, tripartite motif containing 57; TSS, transcription start site; TWIST1, Twist family basic helix-loop-helix transcription factor 1; WIF1, Wnt inhibitory factor 1; WSB1, WD repeat and SOCS box containing 1; YEATS4, YEATS domain containing 4; ZIC1, Zic family member 1; ZNF177, zinc finger protein 177; ZSCAN18, zinc finger and SCAN domain containing 18.

Received September 20, 2019; accepted September 15, 2020.

Additional Supporting Information may be found at onlinelibrary.wiley.com/doi/10.1002/hep.31585/supinfo.

worldwide.^(1,2) GBC is relatively rare in the United States and western Europe but highly prevalent in several countries of South Asia and South America.^(1,3) Because most GBC-related deaths occur in low-income and middle-income countries, research into this disease has long been limited. For years one of the countries with the highest global GBC incidence has been Chile, where GBC is among the leading causes of cancer mortality and morbidity and particularly affects women.^(4,5)

Epidemiological studies have identified several GBC risk factors, including gallstone disease (GSD), inflammation, toxins, ethnicity, and genetic background.^(1,5,6) GSD increases GBC risk 2.4 (gallstones 2.0–2.9 cm in diameter) to 9.2–10.1 fold (gallstones >3 cm).⁽⁷⁾ Due to the high incidence of GBC in Chile, the Chilean government has established a GBC prevention program that relies on prophylactic cholecystectomy for gallstone patients aged 35–49 years.^(4,8) Mechanistically, the association between gallstones

and GBC is suggested to result from the continuous irritation of the gallbladder epithelium, leading to inflammation and enhanced cell regeneration.⁽⁹⁾ This in turn is considered to eventually trigger the progression of epithelial cells through the sequence metaplasia→dysplasia→*in situ* carcinoma, in which cells accumulate genomic and epigenomic alterations that may lead to invasive GBC within 5–15 years.^(1,9–12) An in-depth understanding of the molecular changes that accompany disease progression is essential to improve the early detection and treatment of GBC. However, recent large-scale efforts to characterize the molecular aberrations in GBC did not consider preneoplastic dysplasia lesions or did not investigate epigenome-wide changes in the methylome.^(13,14)

To address this gap, we collected gallbladder tissue specimens from Chilean patients affected by GSD, low-grade dysplasia, high-grade dysplasia, or GBC for methylation profiling with Illumina (San

Supported by the German Federal Ministry of Education and Research (01DN15021), the German Research Foundation (LO 2061/1), the European Union's Horizon 2020 research and innovation program (825741), CONICYT-FONDEF IT16I10050, the Biobank of University of Chile and German Research Foundation (project 314905040-SFB/TRR 209), and the Wilhelm Sander-Stiftung (2015.111.1) (all to S.R.); and by the Else Kröner-Fresenius Stiftung Memorial Grant (2018_EKMS.35) and the Deutsche Krebshilfe through a Mildred Scheel Nachwuchsprogramm Grant (70113307) (both to J.B.). The funders had no role in the design or conduct of the study; the collection, management, analysis, or interpretation of the data; the preparation, review, or approval of the manuscript; or the decision to submit the manuscript for publication.

© 2020 The Authors. HEPATOLOGY published by Wiley Periodicals LLC on behalf of American Association for the Study of Liver Diseases. This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

View this article online at wileyonlinelibrary.com.

DOI 10.1002/hep.31585

Potential conflict of interest: Nothing to report.

ARTICLE INFORMATION:

From the ¹Statistical Genetics Research Group, Institute of Medical Biometry and Informatic, University of Heidelberg, Heidelberg, Germany; ²Molecular Pathology, Institute of Pathology & Department of Translational Genomics, University Hospital of Cologne, Cologne, Germany; ³Mildred Scheel School of Oncology, Medical Faculty, University Hospital Cologne, Cologne, Germany; ⁴Department of Basic and Clinical Oncology, Medical Faculty, University of Chile, Santiago, Chile; ⁵Institute of Pathology, University Hospital Heidelberg, Heidelberg, Germany; ⁶Servicio de Anatomía Patológica, Hospital Clínico de la Universidad de Chile, Santiago, Chile; ⁷Servicio de Anatomía Patológica, Hospital Padre Hurtado, Santiago, Chile; ⁸Facultad de Medicina, Universidad Católica del Maule & Unidad de Anatomía Patológica del Hospital Regional de Talca, Talca, Chile; ⁹Servicio de Anatomía Patológica, Hospital del Salvador, Santiago, Chile; ¹⁰Escuela de Tecnología Médica, Universidad Austral de Chile sede Puerto Montt & Servicio de Anatomía Patológica, Hospital de Puerto Montt, Puerto Montt, Chile; ¹¹Servicio de Anatomía Patológica, Hospital Regional, Arica, Chile; ¹²Servicio de Oncología Médica, Instituto Nacional del Cáncer, Santiago, Chile; ¹³Unidad de Anatomía Patológica, Hospital San Borja Arriaran, Santiago, Chile; ¹⁴Research Unit of Molecular Epidemiology and Institute of Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany; ¹⁵Department of Research, Cancer Registry of Norway, Oslo, Norway; ¹⁶Department of Informatics, University of Oslo, Oslo, Norway; ¹⁷Division of Cancer Epigenomics, German Cancer Research Center, Heidelberg, Germany.

ADDRESS CORRESPONDENCE AND REPRINT REQUESTS TO:

Justo Lorenzo Bermejo, Ph.D.
Statistical Genetics Group
Institute of Medical Biometry and Informatics
University of Heidelberg

Im Neuenheimer Feld 130.3
69120 Heidelberg, Germany
E-mail: lorenzo@imbi.uni-heidelberg.de
Tel.: +496221564180

Diego, CA) Infinium MethylationEPIC BeadChips. These data allowed us to comprehensively examine the methylome and copy-number landscape along the GSD→dysplasia→GBC sequence in Chilean patients. We validated our findings using EpiTYPER MassARRAY technology, investigated the relationship between methylation and gene expression by RNA sequencing (RNAseq) and quantitative RT-PCR, and conducted demethylation experiments in GBC cell lines to functionally assess the effect of methylation on gene expression for several candidate genes.

Materials and Methods

PATIENT COHORT

The study protocol conformed to the ethical guidelines of the 1975 Declaration of Helsinki and was approved by the appropriate ethics committees in Chile. Recruiting centers included seven hospitals across the country, from Arica (north) to Puerto Montt (south); the Chilean Cooperative Group for Oncological Research; and the Biobank of Universidad de Chile. Patients with GSD (those with cholecystectomy but without incipient GBC findings), gallbladder dysplasia, and GBC were invited to participate. With the exception of 2 patients with GBC and missing gallstone information, all patients with GBC and dysplasia investigated in the study carried gallstones. Following written informed consent, the study coordinator interviewed patients and retrieved tissue samples and clinical information using standardized case report forms. We excluded samples stored for >5 years and patients with porcelain gallbladder, polyps, noncholesterol stones, or pancreatic/bile duct abnormalities.

Gallbladder tissue specimens (formalin-fixed, paraffin-embedded [FFPE] $n = 87$, fresh-frozen $n = 2$) were obtained from 88 patients in total (1 patient with GBC contributed one FFPE and one fresh-frozen specimen). For DNA extraction 12 sections per block were cut. The first and last sections were stained with hematoxylin and eosin, and scans can be viewed with QuPath (qupath.github.io) at heidata.uni-heidelberg.de/privateurl.xhtml?token=93181229-aa25-4c53-bcad-cba856ee017b). Morphologically representative regions were chosen after independent review by two experienced pathologists. DNA and RNA were extracted from FFPE sections using the Qiagen AllPrep FFPE

Kit and quality controlled (High Sensitivity Genomic DNA, Advanced Analytical, United States and FFPE quality control kits; Illumina). Fresh-frozen tissue (1–25 mg) was pulverized using a Cole–Parmer EW-36903-00 biopulverizer prior to DNA extraction (QIAamp DNA Micro Kit; Qiagen).

MEASUREMENT OF DNA METHYLATION

Genomic DNA was restored (Infinium FFPE DNA Restoration Kit; Illumina) and bisulfite-converted (EZ-96 DNA Methylation Kit; Zymo Research, Orange, CA) prior to methylation analysis on an Illumina iScan platform using Infinium MethylationEPIC BeadChips containing ~850k methylation markers, according to standard protocols. The processing of raw methylation data and the analysis of differential methylation and copy number variations (CNVs) are described in the Supporting Information.

EpiTYPER MassARRAY METHYLATION ANALYSIS

EpiTYPER MassARRAY (Agena Bioscience) analyses were done as described⁽¹⁵⁾ by treating 1.0 μ g of DNA using the EZ DNA methylation kit (Zymo Research), followed by PCR amplification of regions of interest. Primer design used EpiDesigner software (Agena; Supporting Table S1). Amplicon size was limited to <200 bp to account for short DNA fragments in FFPE DNA. Methylation was validated in all amplicons with DNA methylation standards by variable mixing of methylated and unmethylated DNA generated from human genomic DNA (Roche, Germany) with *in vitro* amplification (RepliG mini kit; Qiagen, Germany) and methylation (M.SssI cytosine–guanine dinucleotide [CpG] methyltransferase; Thermo Fisher Scientific) prior to bisulfite conversion.

GENE EXPRESSION MEASUREMENTS AND IMMUNOHISTOCHEMISTRY ASSAYS

Gene expression was quantified using RNAseq, complemented with quantitative RT-PCR to validate the expression measurements for several candidate genes. RNAseq was performed using the NEBNext Small

RNA Library Prep Set for Illumina (catalog no. E7300; New England Biolabs Inc.) with a cut size on the pip-pin prep (catalog no. CSD3010; Sage Science) covering RNA molecules from 17 to 47 nucleotides, which enables capture of mRNA fragments.⁽¹⁶⁾ Libraries were sequenced on the HiSeq2500 (Illumina) to reach an average depth of 18 million total reads per sample. Total reads were trimmed for adapters using AdapterRemoval v2.1.7⁽¹⁷⁾ and mapped to the human genome (hg38) using Bowtie2 v2.2.9.⁽¹⁸⁾ HTSeq⁽¹⁹⁾ was used to count the reads mapped to mRNA exons in GENCODE v26, applying an established bioinformatics workflow.⁽¹⁶⁾ Read counts were transformed to log₂ transcripts per million. mRNAs with a low count variability (median absolute deviation <2 counts) and samples with fewer than 8 million mapped mRNA reads were excluded from the subsequent statistical analyses. The real-time quantitative PCR expression analyses and immunohistochemistry (IHC) assays conducted for the candidate genes are described in the Supporting Information.

5'-AZA-2'-DEOXYCYTIDINE TREATMENT OF GBC CELL LINES

GBC cell lines OZ (Japanese Cancer Research Resources Bank) and G-415 (RIKEN BioResource Research Center Cell Bank) were regularly tested to be negative for mycoplasma (MycoAlert; Lonza, Basel, Switzerland) and authenticated by short tandem repeat profiling. Cells were cultured at 37°C, 5% CO₂ in Roswell Park Memorial Institute-1640 and OZ cells in Williams' E medium with 10% fetal bovine serum, 1% penicillin-streptomycin (100 IU/mL and 100 g/mL; Sigma-Aldrich, Germany). For demethylation, cells were treated with 1 μM 5'-aza-2'-deoxycytidine (5-aza-dC; Sigma-Aldrich) dissolved in dimethyl sulfoxide (DMSO) or DMSO for 72 hours. Media were changed daily with fresh 5-aza-dC solution. DNA and RNAseq data for G-415 cells were generated as described in the Supporting Information.

Results

MAIN CHARACTERISTICS OF THE CHILEAN COHORT OF PATIENTS WITH GALLBLADDER DISEASE

To investigate molecular changes during the progression from GSD to GBC, we assembled

demographic and clinical data and gallbladder tissue specimens of 88 patients (33 GSD, 15 low-grade dysplasia, 10 high-grade dysplasia, and 30 GBC), of whom 81 fulfilled methylation quality control criteria. We observed higher proportions of females than males (69 women and 12 men), and more older patients were affected by high-grade dysplasia or GBC than by GSD or low-grade dysplasia ($P = 5.1 \times 10^{-5}$; Fig. 1A; Supporting Table S2). The difference in median age between patients with GSD and patients with GBC was 13.5 years, in accordance with the model of carcinogenesis GSD→dysplasia→GBC.⁽¹⁾ Among patients with GBC with available clinicopathological information, only a minority showed advanced tumor (T), node (N), and metastasis (M) stages (25% > T3, 17% > N0, 27% > M0), indicating that the majority of investigated patients with GBC had localized GBC tumors (Supporting Table S2).

GLOBAL METHYLATION DIFFERENCES DURING GBC DEVELOPMENT

For 82 out of 89 investigated samples (92%), the hybridization results from Illumina Infinium MethylationEPIC BeadChips fulfilled the quality control filters when we applied the Minfi package. Principal component analysis (PCA) based on markers with the largest methylation variability revealed similar global methylation patterns for GSD and low-grade dysplasia samples (Fig. 1B) but separated GSD plus low-grade dysplasia from GBC samples. High-grade dysplasia samples fell in between the two groups. Similarly, unsupervised hierarchical clustering discriminated GBC quite accurately from the rest of the samples and correctly classified the two fresh-frozen samples in the GBC group (Fig. 1C).

Bonferroni-Holm-corrected P values from non-parametric Jonckheere-Terpstra tests identified 15,112 markers with monotonically increased or decreased methylation levels along the sequence GSD→low-grade dysplasia→high-grade dysplasia→GBC (Table 1; Supporting Table S3). Interestingly, many of the top 20 markers included CpG sites at or near transcription factors such as zinc finger protein 177 (*ZNF177*) and T-box transcription factor 15 (*TBX15*), including several homeobox genes, for example, BarH-like homeobox 2 (*BARHL2*) and engrailed homeobox 1 (*EN1*). Exemplary box plots with the β methylation

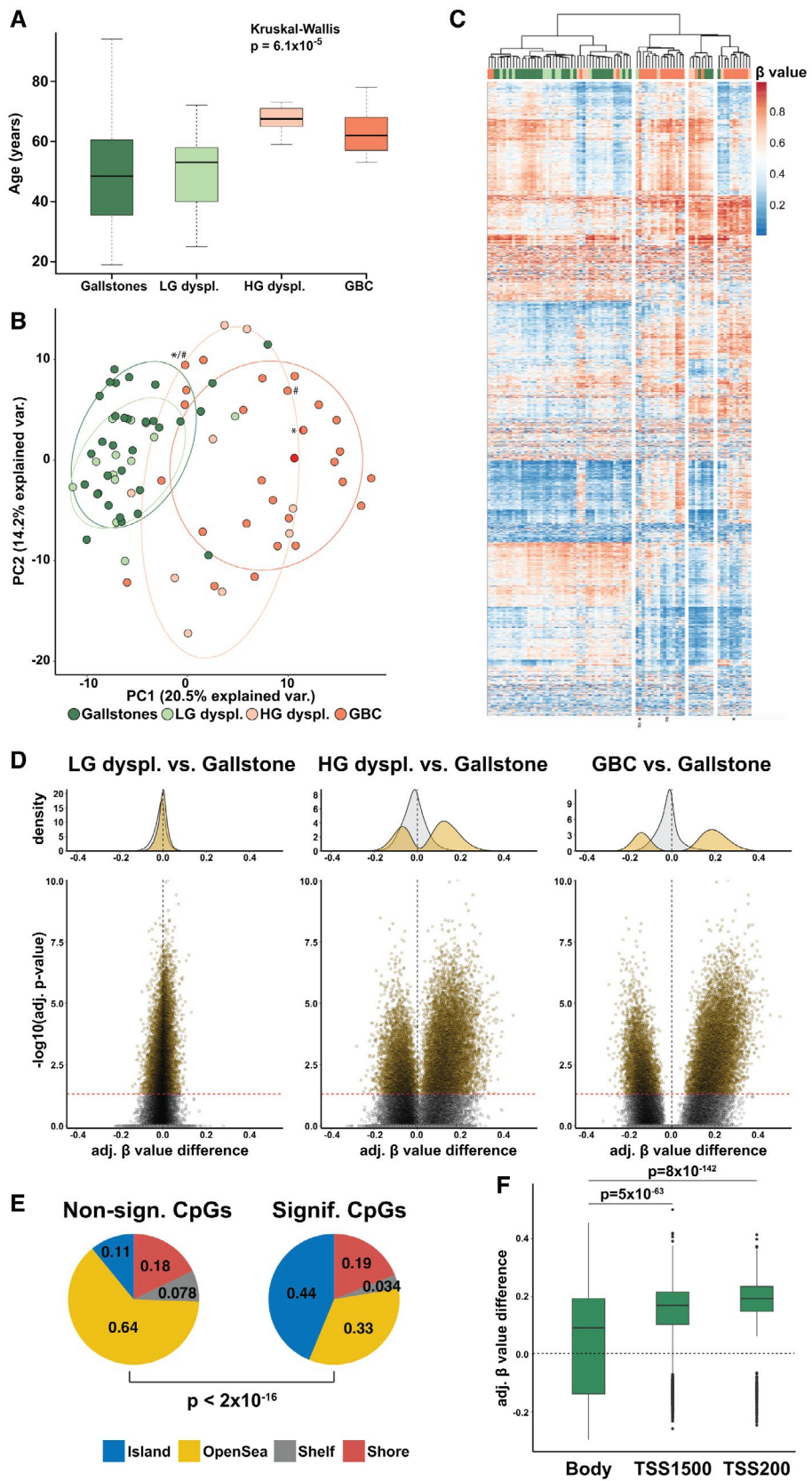


FIG. 1. Global methylation differences during GBC development. (A) Age distribution in the four investigated groups of patients. (B) PCA based on the normalized methylation values for the 10,000 most variable markers (*two samples from the same patient; #two fresh-frozen samples). (C) Heatmap and hierarchical cluster of methylation values for the 10,000 most variable markers (rows), color coding and annotation of samples (columns) according to (A) and (D). (D) Top: Density plots where the *x*-axis shows the average methylation difference compared with GSD samples after adjustment for age and gender using linear regression models. Bottom: Volcano plots for all markers investigated in the differential methylation analysis using Jonckheere-Terpstra tests. The *y*-axis shows $-\log_{10}$ Bonferroni-Holm-corrected Jonckheere-Terpstra *P* values. (E) Distribution of functional genetic elements for differentially (left) and nondifferentially (right) methylated markers (*P* value from Fisher's exact test). (F) Distribution of methylation differences between GSD and GBC samples of significant markers according to their location in gene bodies or 1,500 bp or 200 bp upstream of the nearest TSS (*P* values from two-sample U tests). Abbreviations: HG dyspl., high-grade dysplasia; LG dyspl., low-grade dysplasia; PC, principal component; var. variance.

TABLE 1. Top 20 Markers With Gradual Methylation Changes Along the Sequence GSD → Low-Grade Dysplasia → High-Grade Dysplasia → GBC as Indicated by the Smallest *P* Values From Jonckheere-Terpstra Tests

No.	Methylation Marker	Gene	<i>P</i> *	Mean β Value [†] in GSD Samples	Mean β -Difference [†] LGD vs. GSD	Mean β -Difference [†] HGD vs. GSD	Mean β -Difference [†] GBC vs. GSD
1	cg08493776	PCDHB6	<2e-16	0.04 (0.00;0.14)	0.02 (-0.05;0.10)	0.19 (0.09;0.28)	0.26 (0.19;0.33)
2	cg21392341	TBX15	1.1e-10	0.03 (0.00;0.14)	0.01 (-0.07;0.10)	0.15 (0.04;0.26)	0.21 (0.13;0.28)
3	cg02164046	SST	3.7e-10	0.00 (0.00;0.09)	0.06 (-0.04;0.16)	0.26 (0.13;0.38)	0.36 (0.27;0.44)
4	cg24886257		8.7e-10	0.12 (0.00;0.26)	0.07 (-0.04;0.18)	0.23 (0.09;0.36)	0.35 (0.26;0.45)
5	cg11823511	BARHL2	1.6e-09	0.00 (0.00;0.11)	0.02 (-0.07;0.12)	0.24 (0.12;0.36)	0.29 (0.21;0.38)
6	cg00656990	WVVOX	3.2e-09	0.89 (0.80;0.98)	-0.02 (-0.08;0.05)	-0.17 (-0.26;-0.08)	-0.21 (-0.27;-0.15)
7	cg24503966	NOL4	3.8e-09	0.04 (0.00;0.15)	0.03 (-0.05;0.10)	0.13 (0.03;0.23)	0.23 (0.17;0.30)
8	cg02950416	BCAN	4.3e-09	0.00 (0.00;0.10)	0.03 (-0.05;0.10)	0.14 (0.04;0.24)	0.28 (0.22;0.35)
9	cg26958783	SALL3	4.3e-09	0.10 (0.02;0.17)	0.03 (-0.03;0.08)	0.16 (0.09;0.24)	0.16 (0.11;0.21)
10	cg18359578	KCNMA1	5.5e-09	0.40 (0.35;0.45)	-0.04 (-0.08;0.00)	-0.11 (-0.17;-0.06)	-0.14 (-0.18;-0.10)
11	cg26296488	DRD5	6.2e-09	0.02 (0.00;0.18)	-0.01 (-0.12;0.11)	0.20 (0.05;0.34)	0.36 (0.26;0.46)
12	cg12665460	ZNF578	6.2e-09	0.13 (0.05;0.22)	0.00 (-0.06;0.07)	0.11 (0.03;0.20)	0.21 (0.16;0.27)
13	cg19274890	DPP6	6.2e-09	0.00 (0.00;0.09)	-0.01 (-0.11;0.08)	0.26 (0.13;0.38)	0.37 (0.29;0.46)
14	cg05928342	ZNF177	7.2e-09	0.04 (0.00;0.14)	0.01 (-0.07;0.08)	0.07 (-0.03;0.17)	0.23 (0.16;0.30)
15	cg02519751	ZIC1	7.7e-09	0.08 (0.00;0.22)	0.05 (-0.05;0.15)	0.27 (0.14;0.40)	0.35 (0.26;0.44)
16	cg15885148	CFAP61	9.4e-09	0.83 (0.73;0.92)	-0.05 (-0.12;0.02)	-0.13 (-0.22;-0.04)	-0.20 (-0.26;-0.14)
17	cg03254451	EN1	9.6e-09	0.04 (0.00;0.14)	0.02 (-0.05;0.09)	0.17 (0.08;0.27)	0.23 (0.16;0.29)
18	cg03653841	SFTA3	9.7e-09	0.03 (0.00;0.10)	0.01 (-0.05;0.07)	0.12 (0.04;0.19)	0.17 (0.12;0.22)
19	cg17857974	PCDHGA4	1.1e-08	0.14 (0.06;0.23)	0.05 (-0.02;0.11)	0.15 (0.06;0.23)	0.18 (0.13;0.24)
20	cg14457782	WINK4	1.1e-08	0.08 (0.00;0.17)	0.00 (-0.07;0.07)	0.05 (-0.04;0.14)	0.22 (0.16;0.28)

*Bonferroni-Holm-adjusted *P* value from Jonckheere-Terpstra test.

[†]Average methylation and methylation differences were estimated using a linear model considering the age and gender of the patients as covariates. Mean adjusted β values < 0 in GSD samples were set to 0; if multiple genes are annotated, the first gene is shown.

Abbreviations: BCAN, brevicin; CFAP61, cilia and flagella associated protein 61; DPP6, dipeptidyl peptidase-like 6; DRD5, dopamine receptor D5; HGD, high-grade dysplasia; KCNMA1, potassium/calcium-activated channel subfamily M alpha 1; LGD, low-grade dysplasia; NOL4, nucleolar protein 4; PCDHB6, protocadherin beta 6; PCDHGA4, protocadherin gamma subfamily A, 4; SALL3, SALL1, spalt-like transcription factor 3; SFTA3, surfactant associated 3; SST, somatostatin; WINK4, WNK lysine-deficient protein kinase 4; WVVOX, WW domain containing oxidoreductase.

values for selected markers are shown in Supporting Fig. S1A-D.

In addition to the assessment of significant monotonic changes in methylation, we used linear regression models with the age and gender of the patients as covariates to calculate group-wise methylation levels (Table 1 and Fig. 1D). In accordance with

PCA and cluster analyses, methylation differences between GSD and low-grade dysplasia samples were small; larger differences were observed between GSD and high-grade dysplasia samples, and many markers showed larger differential methylation effects in GBC compared with GSD samples. The majority of differentially methylated markers showed higher

methylation in GBC and high-grade dysplasia than in GSD samples (Fig. 1D).

FUNCTIONAL ANALYSIS OF METHYLATION CHANGES

To gain a deeper understanding of the observed methylation changes during GBC carcinogenesis, we analyzed the functional elements of the identified methylation markers. While most of the 850k markers in the Illumina Infinium MethylationEPIC BeadChip are situated in “open sea” regions, markers with gradual methylation changes during GBC progression were predominantly located in CpG islands (Fisher test $P < 2 \times 10^{-16}$) (Fig. 1E). Markers in gene bodies were underrepresented ($P = 2 \times 10^{-267}$; Supporting Fig. S2A) and those located 1,500 bp or 200 bp within transcription start sites (TSSs; i.e., within promoter regions) were overrepresented ($P = 5 \times 10^{-9}$ and 4×10^{-174} , respectively) among significant markers. On average, markers in gene bodies showed lower methylation than markers in the vicinity of TSSs (TSS1500 $P = 5 \times 10^{-63}$ and TSS200 $P = 8 \times 10^{-142}$), and enhancer methylation decreased with GBC progression (Supporting Fig. S2B).

To gain a systematic overview, we also performed a gene ontology (GO) enrichment analysis based on the neighboring genes of significant markers. GO analysis identified DNA-binding transcription factor activity, membrane organization, and receptor activities as particularly enriched molecular functions (Supporting Fig. S2C). Differential methylation was moreover enriched in pathways involved in receptor signaling, including the Wnt and cadherin pathways, which have been linked to tissue invasion (cell–cell adhesion) in Chilean GBC (Supporting Fig. S2D).^(11,12,20–22) Taken together, these results suggest that the methylation changes during GBC carcinogenesis are functionally relevant and considerably impact genes involved in cell–microenvironment communication.

ANALYSIS OF CANDIDATE GENES

Widespread methylation changes in neighboring CpG sites, particularly those located in gene promoters and CpG islands, tend to be functional. We used the ChAMP R-package, which accounts

for CpG density and methylation changes,⁽²³⁾ to detect differentially methylated regions (DMRs) of clustered CpG sites between high-grade dysplasia plus GBC samples compared with GSD plus low-grade dysplasia samples. Interestingly, the top DMR (Table 2) encompassed promoters of Zic family member 1 (*ZIC1*) and *ZIC4*, two negative regulators of the Hedgehog pathway, previously implicated in GBC pathogenesis.^(24–26) Hypermethylated DMRs were also detected in the promoter regions of two other negative regulators of Hedgehog signaling: hedgehog interacting protein (*HHIP*; 12 CpG sites, mean β log-fold change = 0.11, $P = 3.78 \times 10^{-45}$) and patched 1 (*PTCH1*; 7 CpG sites, mean β log-fold change = 0.04, $P = 5.1 \times 10^{-9}$). Epigenetic regulation of Hedgehog signaling may thus contribute to GBC carcinogenesis in our cohort.

We also detected hypermethylated promoter DMRs in Wnt inhibitory factor 1 (*WIF1*), RUNX family transcription factor 3 (*RUNX3*), and tumor protein P73 (*TP73*) (Fig. 2A), which have been classified as tumor suppressor genes and may therefore be of particular mechanistic relevance to GBC.^(27–29) Figure 2B–D shows the distribution of β values for markers located in promoter regions (gray bars) and gene bodies (orange bars). The heatmaps represent the pairwise correlations of methylation values in patients with GBC, where red blocks point to clustered methylation in neighboring CpG sites. *WIF1* showed a clear, monotonic increase of methylation along the sequence GSD→low-grade dysplasia→high-grade dysplasia→GBC in seven CpG island markers ($P = 8 \times 10^{-41}$; Fig. 2B, top). Moreover, the methylation in the promoter region and the methylation in the body of *WIF1* were negatively correlated (Fig. 2B, middle, blue area), a pattern consistent with transcriptional repression.

RUNX3, which encodes the Runt-related transcription factor 3, showed the most significant promoter hypermethylation among the candidate cancer genes (Fig. 2A). Its main CpG island comprises 21 markers with monotonically increasing methylation during GBC progression ($P = 1.1 \times 10^{-146}$; Fig. 2C, top). Several clusters of coregulated CpGs were identified, including a secondary promoter region lacking a CpG island, which displayed methylation inversely correlated with that of the main CpG island (blue area under second body region to right of Fig. 2C, middle). This complex pattern of

TABLE 2. Regions With the Strongest Methylation Differences Between GSD Plus Low-Grade Dysplasia and High-Grade Plus GBC Samples

Overlapping Promoters	Mean β Log-Fold Change	No. of Markers	Chromosome	Stouffer <i>P</i>
ZIC1, ZIC4	0.20	62	3	0
DMRTA2	0.18	54	1	0
ZIC1, ZIC4, ZIC4-AS1	0.21	42	3	0
PAX3, CCDC140	0.17	54	2	8.01E-301
IRX2, C5orf38	0.18	40	5	6.45E-290
FEZF2, PTPRG-AS1	0.14	56	3	9.25E-288
IRX4, CTD-2194D22.3	0.17	48	5	1.93E-286
SALL1	0.18	43	16	7.61E-282
HAND2	0.18	39	4	2.23E-280
WT1	0.17	46	11	9.82E-269
EDNRB, RNF219-AS1	0.18	39	13	1.82E-265
SOX14	0.17	39	3	1.61E-256
NKX2-1	0.19	34	14	9.39E-254
SIX6	0.19	29	14	1.20E-242
DLX5	0.15	51	7	9.64E-242

Abbreviations: AS1, antisense 1; CCDC140, CCDC140 long noncoding RNA; C5orf38, chromosome 5 open reading frame 38; DLX5, distal-less homeobox 5; DMRTA2, DMRT-like family A2; IRX2, Iroquois homeobox 2; IRX4, Iroquois homeobox 4 (CTD-2194D22.3); NKX2-1, NK2 homeobox 1; PAX3, paired box 3; PTPRG, protein tyrosine phosphatase receptor type G; RNF219, RING finger protein 219; SALL1, spalt-like transcription factor 1; SIX6, SIX homeobox 6; SOX14, SRY box transcription factor 14; WT1, WT1 transcription factor.

epigenetic regulation corroborates the advantage of high-density methylation arrays over single-marker analyses.

The complexity of methylation profiles was even more evident for *TP73*, which comprises two promoter regions that may favor the transcription of tumor-promoting or tumor-inhibiting isoforms.⁽²⁹⁾ We detected a hypermethylated region in the CpG island of the promoter of *TP73*, upstream of the TSS and outside the coding region (12 markers, $P = 2 \times 10^{-56}$; Fig. 2D), which may favor the expression of the long tumor-inhibiting isoform.⁽²⁹⁾ In addition to monotonic methylation changes during GBC development, we assessed potential methylation shifts in GBC samples according to the tumor characteristics. In line with functional experiments showing an inhibitory effect of *WIF1* not only on proliferation but also on migration and invasion of GBC cancer cells,⁽²⁰⁾ the methylation of *WIF1*, *RUNX3*, and *TP73* increased with tumor grade (Fig. 2E) and stage (Supporting Fig. S3A); and *WIF1* showed higher methylation in metastatic GBC (Supporting Fig. S3B).

We also detected DMRs in the promoter regions of other candidate cancer genes (Fig. 2A), including genes found to be differentially methylated in cholangiocarcinoma (CCA) and GBC.^(10,12,30) For example, clusters of

hypermethylated CpG sites were found in tumor protein p53-dependent G₂ arrest mediator homolog (*RPRM*) and Twist family basic helix-loop-helix transcription factor 1 (*TWIST1*), and hemoglobin subunit epsilon 1 (*HBE1*) presented two hypomethylated neighboring CpGs. The regional hypermethylation of *TWIST1* and the hypomethylation of *HBE1* were associated with tumor characteristics of the GBC samples (Supporting Fig. S3C-E). Hypermethylated promoter regions were also noticed in doublecortin-like kinase 1 (*DCLK1*), cysteine dioxygenase type 1 (*CDO1*), *ZNF331*, and zinc finger and SCAN domain containing 18 (*ZSCAN18*), frequently found to be hypermethylated in other gastrointestinal tumors, including colorectal cancer.⁽³⁰⁾ Ribosomal protein L22 (*RPL22*), recently categorized as a potential tumor suppressor in CCA, showed six neighboring hypermethylated markers (Fig. 2A; Supporting Fig. S3F).⁽³¹⁾ *CDO1*, *ZNF177*, branched chain amino acid transaminase 1 (*BCAT1*), and tripartite motif containing 57 (*TRIM57*), which have been proposed as a four-gene methylation-based diagnostic biomarker for non-small cell lung cancer, also showed hypermethylated promoter DMRs (Fig. 2A).⁽³²⁾ In addition to the widespread hypermethylation in GBC, we identified genes with promoter hypomethylation and protumorigenic properties reported for other cancer entities. They

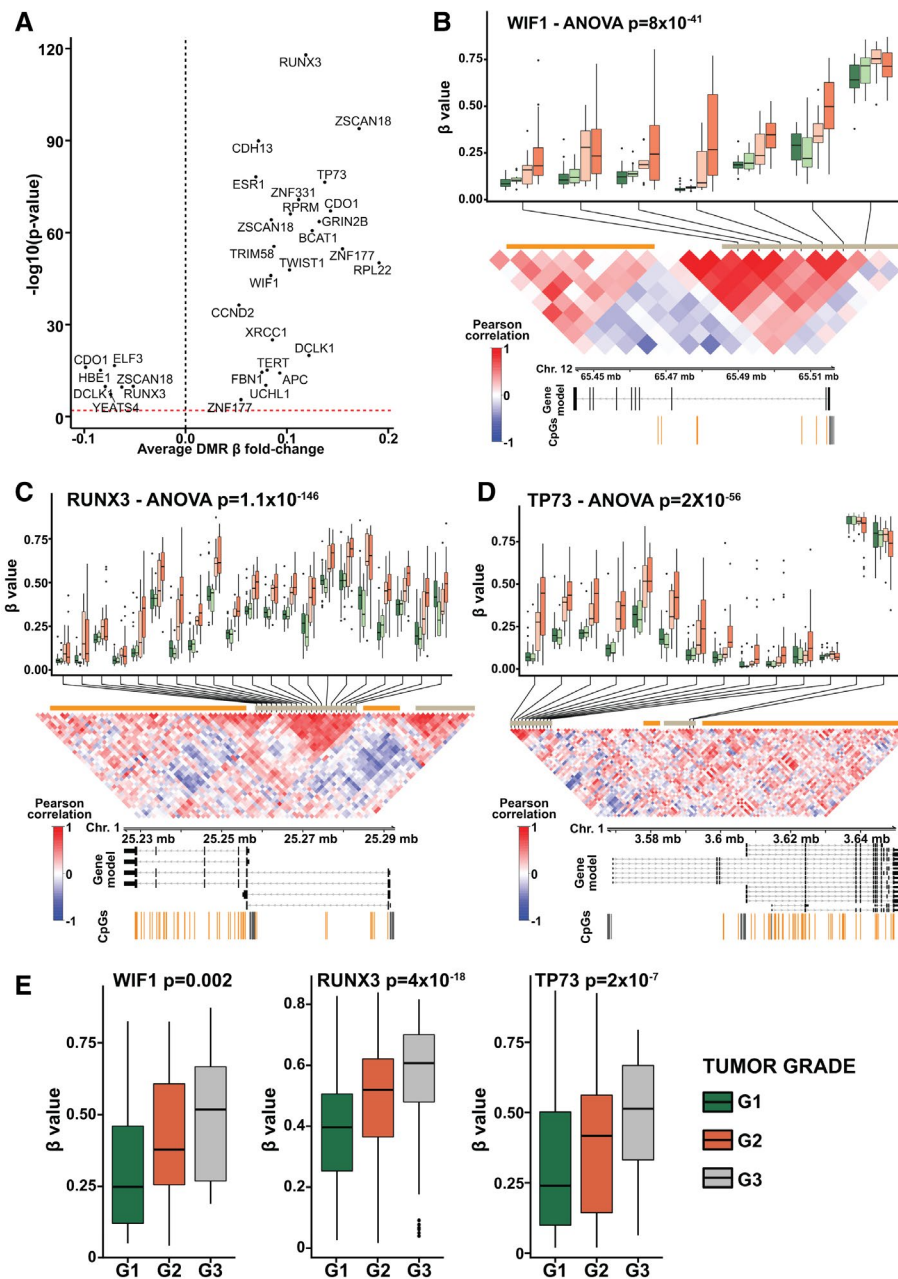


FIG. 2. Differential methylation of candidate genes. (A) Candidate cancer genes with DMRs in their promoters and average methylation fold changes between GSD plus low-grade dysplasia and high-grade dysplasia plus GBC samples. (B-D) Top: Box plots of methylation values for the four investigated groups of patients in the promoter-associated CpG islands of *WIF1*, *RUNX3*, and *TP73*, respectively (P values from analysis of variance [ANOVA] tests adjusting for baseline CpG differences: green, GSD; light green, low-grade dysplasia; light red, high-grade dysplasia; red, GBC). Middle: Heatmaps show the pairwise Pearson correlation of methylation values between marker pairs (gray bar, promoter region; orange bar, gene body). Bottom: Annotation of the CpGs with genomic location and position along the genes' isoforms (gray, CpGs in promoter region; orange, CpGs in gene body). (E) Distributions of methylation values in the promoter CpG islands of *WIF1*, *RUNX3*, and *TP73* according to tumor grade (P values from ANOVA tests adjusting for baseline CpG differences). Abbreviations: CDH13, cadherin 13; ELF3, E74-like ETS transcription factor 3; ESR1, estrogen receptor 1; GRIN2B, glutamate ionotropic receptor *N*-methyl-D-aspartic acid type subunit 2B; TERT, telomerase reverse transcriptase; UCHL1, ubiquitin C-terminal hydrolase L1; XRCC1, X-ray repair cross-complementing protein 1.

included high-mobility group AT-hook 1 (HMGA1)⁽³³⁾ (mean β log-fold change -0.11 , $P = 5 \times 10^{-22}$), Erb-B2 receptor tyrosine kinase 2 (ERBB2) (mean β log-fold change -0.035 , $P = 7 \times 10^{-4}$), cell division cycle associated 7 (CDCA7)⁽³⁴⁾ (mean β log-fold change -0.11 , $P = 3 \times 10^{-27}$), ankyrin 1 (ANK1)⁽³⁵⁾ (mean β log-fold change -0.10 , $P = 2 \times 10^{-39}$), RUNX1⁽³⁶⁾ (mean β log-fold change -0.13 , $P = 1 \times 10^{-28}$), WD repeat and SOCS box containing 1 (WSB1)⁽³⁷⁾ (mean β log-fold change -0.105 , $P = 2 \times 10^{-8}$), and YEATS domain containing 4 (YEATS4)⁽³⁸⁾ (mean β log-fold change -0.07 , $P = 5 \times 10^{-8}$). Our analyses thereby detected genes with a potential tumor-promoting function in GBC, which constitute bona fide candidates for experimental follow-up studies. All of these results highlight potential mechanistic and epigenetic similarities among different cancer entities.

Sensitivity analyses using the GenomeStudio instead of the Minfi package for quality control of methylation data resulted in 81 of 89 samples (91%) passing quality control filters, leading to highly correlated P values ($r = 0.98$) and estimated methylation differences ($r = 0.80$) and to largely similar DMR-volcano plots (Supporting Fig. S4A-C).

Interestingly, we identified methylation changes according to tumor stage and grade in several candidate genes (Fig. 2E; Supporting Fig. S3), but the comparison of G1 versus G2+G3 GBC tumors and of high-grade dysplasia versus GBC and the investigation of the sequence high-grade dysplasia \rightarrow G1 \rightarrow G2 \rightarrow G3 using Jonckheere-Terpstra tests identified neither DMRs nor differentially methylated CpGs after correction for multiple testing. This finding is consistent with the overlap between high-grade dysplasia and GBC samples in the PCA (Fig. 1B) and indicates that larger studies are still needed to investigate the grade-related epigenetic alterations in GBC.

COPY NUMBER ANALYSIS

Infinium MethylationEPIC arrays offer the opportunity to investigate genome-wide CNVs. Along the sequence GSD \rightarrow low-grade dysplasia \rightarrow high-grade dysplasia \rightarrow GBC the frequency of altered copy number segments per sample increased ($P = 6.8 \times 10^{-8}$; Fig. 3A). This suggests that the genomic instability of high-grade dysplasia and GBC is higher than that of low-grade dysplasia and gallstone samples, leading to more

genomic abnormalities (Fig. 3B-F). Among the genes most frequently affected were the tumor suppressors cyclin-dependent kinase inhibitor 2A (*CDKN2A*) and *TP53*, with genomic losses in 8 (29%) and 5 (18%) GBC samples, respectively (Fig. 3B, D). We found frequent copy number gains of *MDM2* proto-oncogene (*MDM2*; GBC, 21%; high-grade dysplasia, 22%), a major negative regulator of *TP53* and of cyclin D1 (*CCND1*) which drives cell cycle progression (Fig. 3B, C). *YEATS4*, described as another negative regulator of the *TP53* pathway and an oncogene in lung cancer,⁽³⁸⁾ was hypomethylated in GBC samples, as described. The similar CNV profiles for *MDM2* and *YEATS4* may suggest alternative mechanisms of functional *TP53* silencing not yet described in GBC, warranting future experimental research.

The simultaneous analysis of the frequency and the magnitude of CNVs using the software GISTIC2⁽³⁹⁾ revealed recurrent alterations in *MDM2*, *YEATS4*, *CCND1*, and *CDKN2A* (Supporting Fig. S5), vindicating the single-sample results. We detected a recurrent (14%) amplification at 17q21.2 in close proximity to the *ERBB2* locus (17q12). This amplification peak comprised keratin genes not previously involved in GBC, the gastrin precursor gene, eukaryotic translation initiation factor 1, and adenosine triphosphate citrate lyase, a metabolic enzyme investigated as a therapeutic target.⁽⁴⁰⁾ Further genomic analyses of Chilean GBC samples and the functional assessment of these candidate genes may thus enhance our knowledge of the mechanisms of GBC to identify therapeutic strategies.

Considering the possibility of personalized treatment with kinase inhibitors, kinase genes affected by CNVs included *ERBB2* (Her2; gains in 14% of GBC samples); cyclin-dependent kinase 4 (*CDK4*; gains in 22% of high-grade dysplasia samples); fibroblast growth factor receptor 3 (*FGFR3*); *ERBB3*; MET proto-oncogene, receptor tyrosine kinase (*MET*); and *CDK6* (Fig. 3B). The fibroblast growth factor ligands *FGF3*, *FGF4*, and *FGF19* were coamplified in the four (14%) GBC samples with *CCND1* gains (Fig. 3B; Supporting Fig. S5), suggesting that tailored combination therapy could be indicated in these patients with GBC. Taken together, the CNV results portray the landscape of copy number alterations in Chilean patients and point to genes that could be targeted by cancer drugs following further mechanistic analysis.

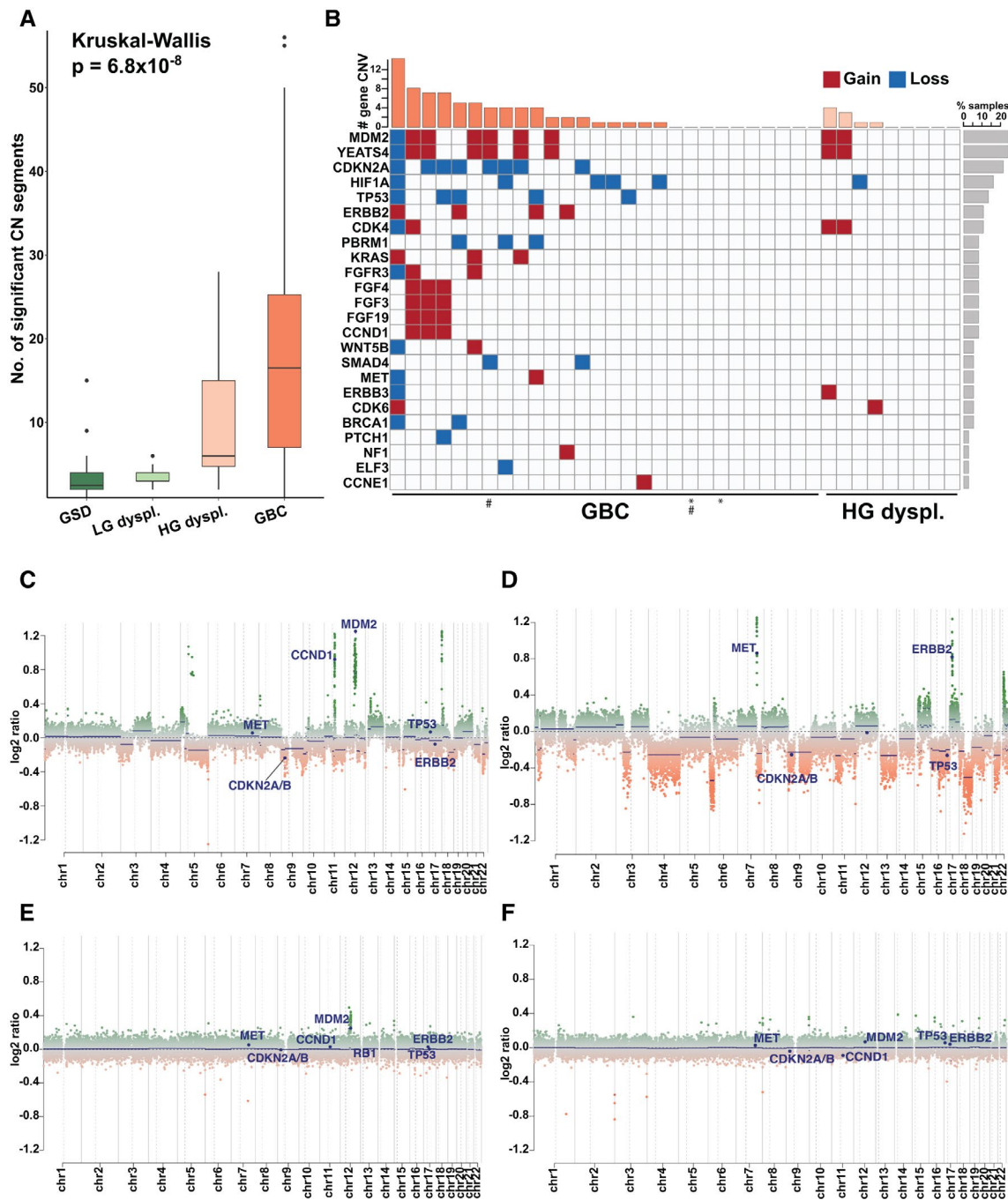


FIG. 3. Copy number profiles in the investigated Chilean cohort. (A) Number of significant copy number segments in the four investigated groups of patients. (B) CNVs for candidate genes in GBC and high-grade dysplasia samples (red, genomic gain; blue, genomic loss). The bar graph at the top shows the number of altered candidate genes per sample, while the bar graph to the right indicates the percentage of altered samples for the respective gene (*two samples from the same patient; # two fresh-frozen samples). (C-F) Typical genome-wide copy number plots: (C) GBC sample with gains in *MDM2* and *CCND1*; (D) GBC sample with gains in *MET* and *ERBB2*; (E) high-grade dysplasia sample; (F) GSD sample. Abbreviations: BRCA1, breast cancer type 1; CCNE1, cyclin E1; CN, copy number; ELF3, E74-like ETS transcription factor 3; HG dyspl., high-grade dysplasia; HIF1A, hypoxia inducible factor 1 subunit alpha; KRAS, KRAS proto-oncogene, guanosine triphosphatase; LG dyspl., low-grade dysplasia; NF1, neurofibromin 1; PBRM1, polybromo 1; SMAD4, mothers against decapentaplegic homolog 4; WNT5B, Wnt family member 5B.

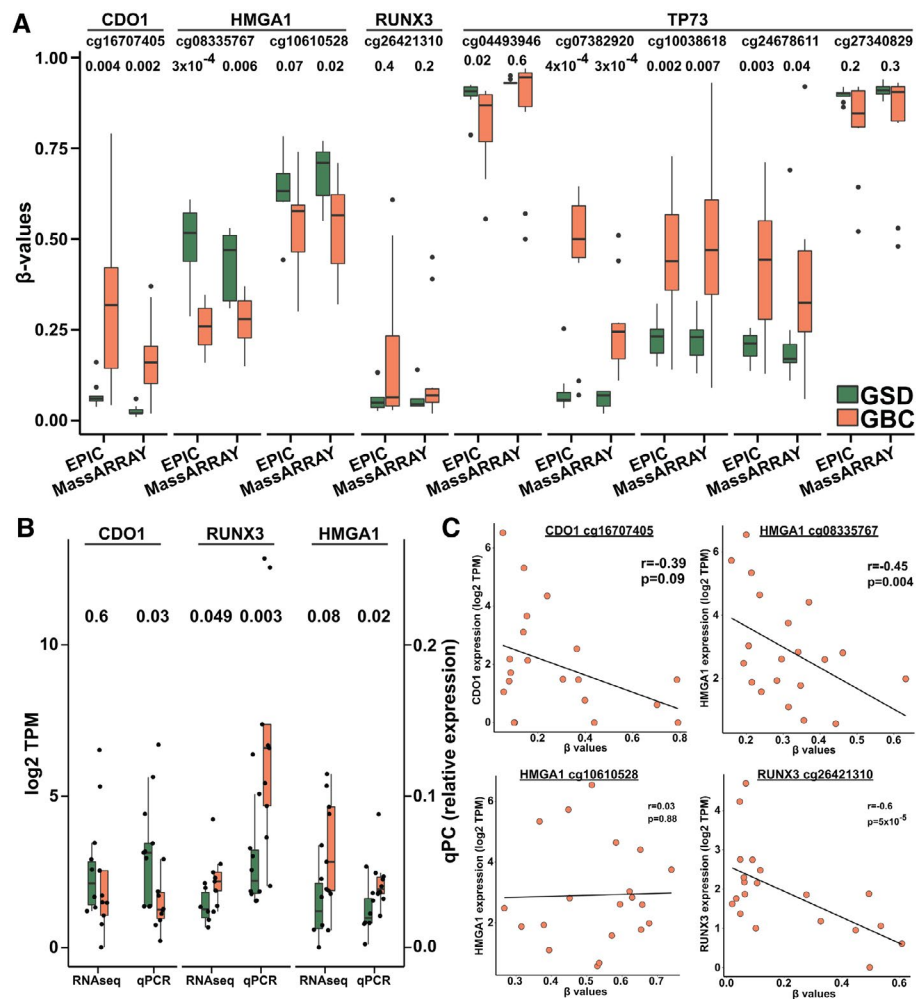


FIG. 4. Validation of methylation differences and relationship between DNA methylation and RNA expression. (A) Box plots with β -methylation values in samples from patients with GSD ($n = 10$) and patients with GBC ($n = 10$) for selected CpGs. Methylation was quantified using epigenome-wide EPIC and candidate-marker MassARRAY technology. The P values in the upper part of the panel are based on two-sample U tests (GSD versus GBC). (B) Gene expression levels determined by RNAseq (log₂ transcripts per million; left y -axis) and quantitative RT-PCR (cycle threshold values normalized to two housekeeping genes to quantify the relative expression in arbitrary units; right y -axis) for GSD ($n = 8$ -10) and GBC ($n = 9$ -10) validation samples. P values are based on two-sample U tests. (C) Correlation between β values for the CpGs shown in (A) and mRNA expression in GBC samples with available RNAseq mRNA and EPIC methylation data ($n = 20$). Pearson correlation coefficients and the corresponding P values are shown. Abbreviations: qPC/qPCR, quantitative RT-PCR; TPM, transcripts per million.

VALIDATION OF METHYLATION MEASUREMENTS AND INVESTIGATION OF THE RELATIONSHIP BETWEEN METHYLATION AND mRNA EXPRESSION

To validate the identified methylation differences using Illumina Infinium MethylationEPIC

BeadChips, we applied the PCR-based EpiTYPER MassARRAY technology and quantified the methylation of nine candidate CpG markers in a subset of the available DNA samples (GSD, $n = 10$; GBC, $n = 10$). The side-by-side plots in Fig. 4A showed good overall consistency between the results based on epigenome-wide EPIC and candidate-marker MassARRAY measurements ($n = 19$ matched samples; Supporting Table S4).

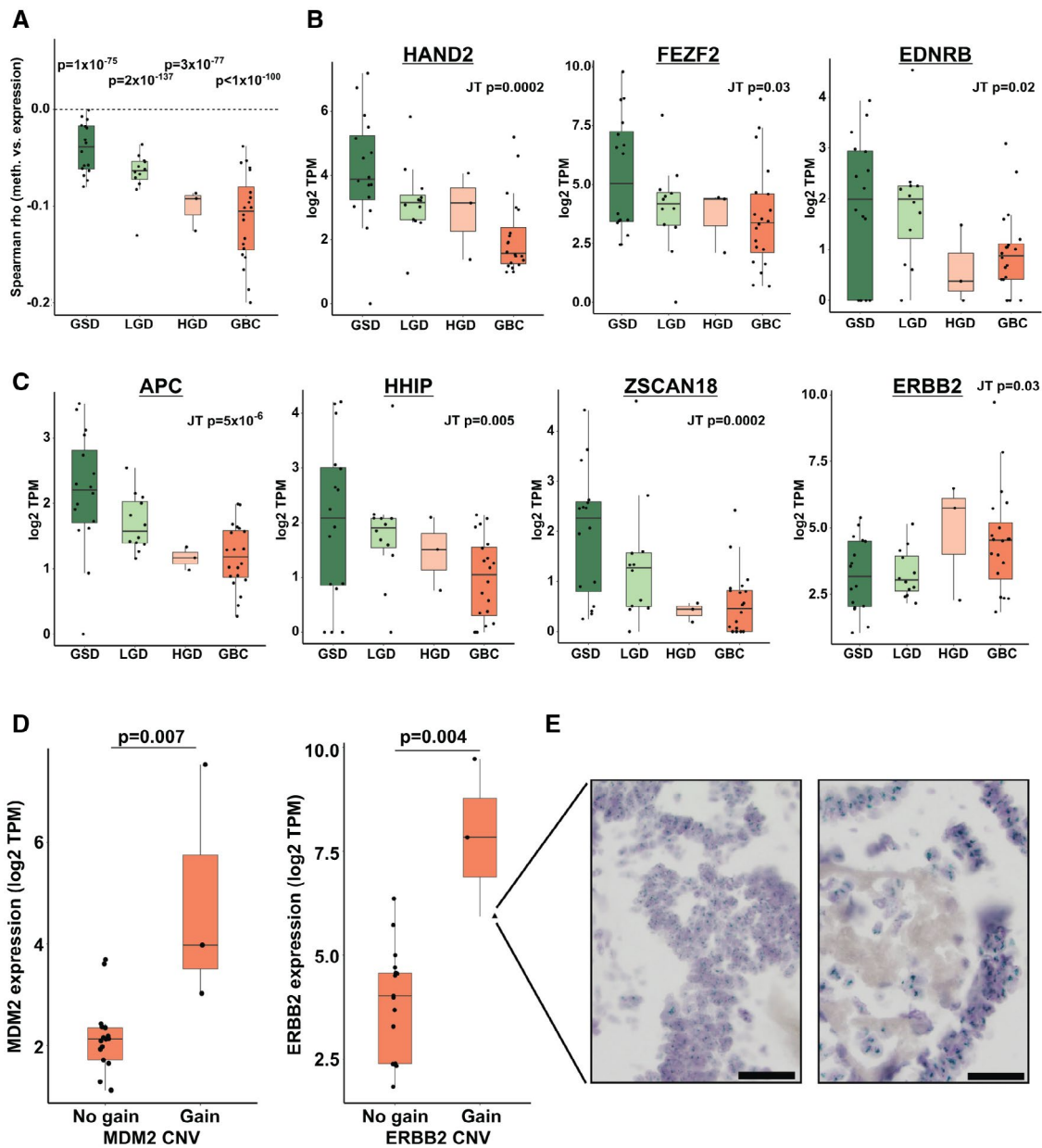


FIG. 5. Expression analysis of candidate genes. (A) Spearman correlation coefficient between the median promoter EPIC methylation and RNAseq mRNA expression. The individual correlation P values for each sample were combined using Fisher's method. (B) RNAseq mRNA expression of differentially methylated genes shown in Table 2 (P values from Jonckheere-Terpstra tests across disease groups). (C) RNAseq mRNA expression of candidate genes showing hypermethylated (*APC*, *HHIP*, *ZSCAN18*) or hypomethylated DMRs (*ERBB2*). P values from Jonckheere-Terpstra tests across disease groups. (D) *MDM2* and *ERBB2* RNAseq mRNA expression in GBC samples without and with copy number gains in *MDM2* and *ERBB2*. P values are based on two-sample U tests. (E) Validation of *ERBB2* copy number gains in two GBC samples using *ERBB2* dc-CISH. The left picture was obtained from the sample highlighted with a triangle in the plot to the right of (D). The right picture is from a second patient with a predicted *ERBB2* gain. Scale bar, 20 μ m. Abbreviations: HGD, high-grade dysplasia; JT, Jonckheere-Terpstra; LGD, low-grade dysplasia; TPM, transcripts per million.

We next applied small RNAseq, which allowed us to quantify the mRNA expression of around 12,000 genes per sample on average (Supporting Fig. S6A), complemented with quantitative RT-PCR to validate

the expression differences between GSD and GBC samples for the candidate genes *CDO1*, *RUNX3*, and *HMGAI1*. The side-by-side plots in Fig. 4B revealed a good overall consistency between RNAseq and

quantitative PCR results. We noticed a negative correlation between RNAseq expression and EPIC methylation for *CDO1*, *HMGAI* (cg08335767), and *RUNX3* in patients with GBC (n = 20; Fig. 4C) as well as in the complete set of investigated GSD, dysplasia, and GBC samples (n = 51; Supporting Fig. S6B). In agreement with RNAseq and quantitative PCR results, IHC analysis of the available slides (n = 11) showed a higher *CDO1* expression in GSD than in GBC ($P = 0.01$; Supporting Table S5) and a negative relationship between *CDO1* expression and methylation ($P = 0.02$; Supporting Fig. S7A,B). However, IHC for *CDO1* in an independent group of 10 Chilean patients with GBC with paired tumor and adjacent nontumor samples did not show differences in expression (increased expression in n = 3, equal expression in n = 4, and decreased expression in n = 3 pairs; Supporting Fig. S7C).

Within each disease stage, we found a negative correlation between the median promoter methylation and mRNA expression, and the correlation strength increased along the sequence GSD→low-grade dysplasia→high-grade dysplasia→GBC (Fig. 5A). Accordingly, the three genes heart and neural crest derivatives expressed 2 (*HAND2*), FEZ family zinc finger 2 (*FEZF2*), and endothelin receptor type B (*EDNRB*), hypermethylated in our DMR analyses (Table 2) and with available RNAseq mRNA expression data, showed decreasing expression levels toward GBC (Fig. 5B). We also observed decreasing expression values for the hypermethylated genes APC regulator of WNT signaling pathway (*APC*; Wnt signaling), *HHIP* (Hedgehog signaling), and *ZSCAN18* (Fig. 5C), while the hypomethylated genes *HMGAI* and *ERBB2* showed increasing expression levels (Figs. 4B and 5C, respectively). Moreover, several genes with gradual methylation changes along the progression from GSD to GBC according to Jonckheere-Terpstra tests (Table 1) showed the contrary tendencies in mRNA expression (Supporting Table S6), further indicating a functional relevance of the observed methylation changes. The previously reported hypermethylation and decreased expression of fibrillin 1 (*FBN1*), superoxide dismutase 3, and LIM domain containing preferred translocation partner in lipoma in Indian patients with GBC was also apparent in our cohort of Chilean patients with GBC⁽¹⁰⁾ (Supporting Fig. 6C).

In accordance with the EPIC array-based CNV data, *MDM2* and *ERBB2* were overexpressed in

samples that showed copy number gains (Fig. 5D). The validation of the detected *ERBB2* gains in 2 patients using *ERBB2* dual-color chromogenic *in situ* hybridization (dc-CISH) lends additional support to our results (Fig. 5E).⁽⁴¹⁾

RESULTS OF DEMETHYLATION ASSAYS

To further investigate potential causal effects of methylation on gene expression, we conducted functional analyses in GBC cell lines. Figure 6A depicts the relationship between epigenome-wide DNA methylation and RNAseq data for the GBC cell line G-415. Overall, methylated promoter regions show low expression values including hypermethylated candidate genes such as *CDO1*, *TP73*, *RUNX3*, *WIF1*, *TRIM58*, *ZNF177*, and *ZSCAN18* (Fig. 6A). By contrast, hypomethylated genes in GBC with reported protumorigenic effects (e.g., *YEATS4*, *WSB1*, *CDCA7*, *ANK1*, and *HMGAI*) showed in general low promoter methylation and higher expression in G-415 cells.

To functionally assess the effect of methylation on gene expression for some candidate genes, we treated two GBC cell lines (G-415 and OZ) with the DNA-methyltransferase inhibitor 5-aza-dC. Figure 6B shows the β -methylation values, and Fig. 6C depicts the expression levels of *CDO1*, *RUNX3*, *TP73*, and *HMGAI* after 72-hour treatment. Compared to controls, 5-aza-dC treatment reduced the methylation of *CDO1*, *RUNX3*, and *TP73* in both cell lines, while the methylation of *HMGAI* (hypomethylated in GBC) was unaltered (Fig. 6B). Accordingly, 5-aza-dC treatment resulted in increased expression of *CDO1* and *RUNX3* in both cell lines, supporting a transcriptional derepression by decreased DNA methylation (Fig. 6C). The lower methylation of *TP73* after treatment translated into increased expression in G-415 cells, but no expression change was noticed in OZ cells. As expected, the changes of *HMGAI* expression were not statistically significant (Supporting Table S4).

Discussion

In the present study we report on epigenome-wide changes in methylation and CNVs along the sequence GSD→low-grade dysplasia→high-grade

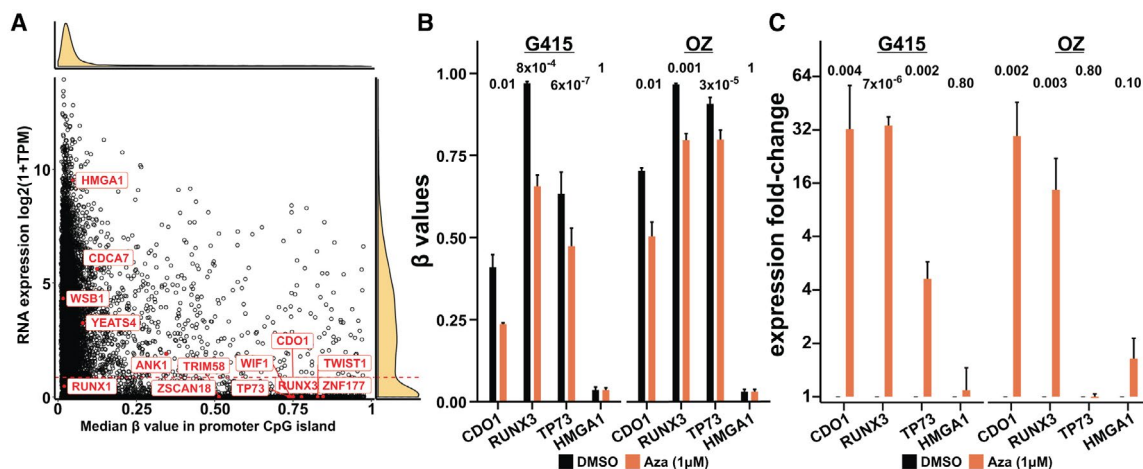


FIG. 6. Functional validation of methylation and RNA expression in GBC cell lines. (A) Scatterplot of median methylation β values of CpGs in promoter CpG islands versus RNA expression quantified as transcripts per million by RNAseq for the GBC cell line G415. Several candidate genes in the present study are highlighted. Density plots show the global distribution of methylation and expression values. (B) Average methylation of the same CpGs investigated in (A) for the GBC cell lines G-415 and OZ after 72-hour treatment with 5-aza-dC or DMSO as negative control ($n = 3$). Displayed are mean + SEM. The P values in the upper part of the panel are based on ANOVA considering the baseline CpG values and multiple CpGs per gene if measured. (C) Gene expression levels determined by quantitative RT-PCR for the GBC cell lines G-415 and OZ after 72-hour treatment with 5-aza-dC or DMSO as negative control ($n = 3$). Cycle threshold values were normalized to three housekeeping genes, and fold changes were calculated using the $2^{-\Delta\Delta C_t}$ method. Displayed are mean \pm SEM. The P values in the upper part of the panel are based on t tests. Abbreviations: Aza, 5-aza-dC; TPM, transcripts per million.

dysplasia \rightarrow GBC in Chilean patients. Chile is possibly the country with the highest prevalence of GBC worldwide, and GBC is the second most frequently occurring cancer-related cause of death among Chilean women.^(4,5) Previous studies investigated Indian patients, who show lower rates of GSD (86% for Chilean compared with 33% for Indian patients with GBC),⁽⁴²⁾ and applied a different study design (fresh-frozen tumor, adjacent nontumor, and gallstone samples).^(10,43) An enhanced understanding of the molecular alterations occurring during GBC development in regions with a high incidence of both GSD and GBC, such as southern Chile, is crucial to improve GBC prevention and treatment in order to ultimately reduce the GBC burden.

One of the most established GBC risk factors in Amerindian and European populations is GSD, which is considered to induce GBC through inflammation following a dysplasia \rightarrow carcinoma sequence.^(1,5,6,9,12) This holds true especially in Chile and was reflected in a previous gallstone history for all clinically annotated GBC samples in our cohort. However, longitudinal samples from the same patient are practically unattainable,

and direct evidence of GSD \rightarrow dysplasia \rightarrow GBC development is scarce. We therefore designed this study to include samples from GSD through dysplasia to GBC. In our cohort, the observed age difference of 13.5 years between patients with GSD and patients with GBC is in accordance with the time postulated for transformation of GSD to GBC.^(1,9,11,12) Similarly, unsupervised PCA clustered GBC samples furthest away from GSD samples, with high-grade dysplasia samples situated in between, suggesting a sequential process. Along this sequence we observed substantial epigenomic changes with a general methylation increase. Hypermethylation predominantly affected CpG islands and promoters, which play essential roles in transcriptional control and may thus rewire gene expression programs during GBC development.

Interestingly, whereas high-grade dysplasia and GBC were clearly distinguishable from benign samples on the epigenetic level, only minimal methylation differences were noticed between GSD and low-grade dysplasia samples. We cannot rule out the possibility that detection of initial molecular changes was hampered by the smaller fraction of dysplastic cells in

low-grade than in high-grade dysplasia samples or by differences in the demographic and clinical characteristics of the patients. The similar distributions of patient's age and CNVs in gallstone and low-grade dysplasia, however, further corroborated systematic differences from high-grade dysplasia and GBC. The minor differences identified may therefore indicate that early GBC stages possess limited molecular alterations, which, moreover, could manifest primarily at the mutational or transcriptomic rather than the epigenomic level. Recent large-scale projects have focused mainly on the comparison of benign versus cancer samples and thus lack intermediary premalignant stages.^(13,14,44) Additional studies specifically addressing genomic changes along the dysplasia→carcinoma sequence will therefore be essential to clarify early molecular drivers of GBC development.

We identified several candidates that offer potential insights into mechanistic regulation including promoter hypermethylation of tumor suppressors *RUNX3* and *TP73* and hypomethylation of *HMGA1*.^(27,28,31,33) *RUNX3* was recently shown to be progressively hypermethylated in hepatocytes during aflatoxin-induced malignant transformation.⁽⁴⁵⁾ Of note, aflatoxin:albumin adducts were found more often in Chilean patients with GBC than in healthy community controls, suggesting greater aflatoxin exposure in patients with GBC.⁽⁴⁶⁾ Unfortunately, no information on aflatoxin exposure was available for our cohort. Further investigations of the role of aflatoxin as a potentially causal risk factor in Chilean patients seem warranted in order to improve prevention of GBC. Interestingly, the expression and the DNA methylation of *RUNX3* were negatively correlated in GBC samples, but *RUNX3* showed increased methylation (quantified both by EPIC and by MassArray) and simultaneously increased expression (according to RNAseq and quantitative PCR) in GBC compared to GSD samples. This finding suggests potential heterogeneity and a complex role of *RUNX3* regulation in GBC.

We also identified and validated alterations in the Wnt and Hedgehog signaling pathways, previously implicated in GBC, CCA, and other tumor types.^(24-26,47) The detected alterations included hypermethylation of *APC*, *WIF1*, and *HHIP* and underexpression of *HHIP* and *APC* in GBC. The available RNAseq data for *WIF1* did not surpass the quality control filters, but previous functional studies have shown a suppressive effect of *WIF1* on Wnt signaling, cell

proliferation, and invasion.^(20,27) In our GBC samples, *WIF1* methylation increased with T stage, and its epigenetic silencing may thus be a mechanism to increase malignant GBC properties. Several inhibitors targeting Wnt and Hedgehog signaling have recently been tested in clinical trials.^(48,49) An improved knowledge of these pathways may thus not only foster the mechanistic understanding of GBC pathogenesis but also lead to GBC treatment strategies.

In addition to epigenetic information, methylation arrays offer the opportunity to investigate copy number changes.^(50,51) Interestingly, we observed not only frequent loss of *TP53* but also co-occurring gains of the *TP53*-inactivating genes *MDM2* and *YEATS4*⁽³⁸⁾ exclusively in patients without *TP53* loss. This suggests alternative modes of *TP53* inactivation, even though the role of *YEATS4* in GBC is currently unknown. Unexpectedly, we did not find any genomic loss of *TP53* in high-grade dysplasia samples, described as an early GBC event.⁽¹²⁾ This could be due to our small number of high-grade dysplasia samples (n = 10) or the fact that our study design does not allow the analysis of mutations, a major driver of *TP53* inactivation. The observed increase of *MDM2* and *ERBB2* expression and the validation of *ERBB2* gain using dc-CISH in two GBC samples with a predicted copy number gain add plausibility to our findings.

A concomitant factor for the dismal prognosis of GBC is the lack of molecular therapeutic targets. We therefore assessed alterations in targetable genes and identified genomic gains of *ERBB2*, *ERBB3*, *MET*, *FGFR3*, *CDK4*, and *CDK6*. Whether the detected genomic alterations constitute driver events in GBC requires investigation in larger cohorts and preclinical GBC models. However, inhibitors targeting the corresponding proteins are clinically used for other tumor entities and may provide personalized treatment options for GBC. A recent study in a Chinese GBC cohort identified *ERBB2* mutations that promoted immune evasion, leading to preclinical activity of targeted *ERBB* kinase inhibition combined with immune checkpoint blockade.⁽⁵²⁾ These results highlight the potential of kinase inhibition in the treatment of GBC and offer additional avenues to improve GBC therapy.

In conclusion, we assessed epigenome-wide changes in methylation and CNVs along the sequence GSD→low-grade dysplasia→high-grade dysplasia→GBC. By investigating a Chilean cohort of

patients we provide insights into the pathways involved in GBC pathogenesis within this specific geographic and genetic environment and provide potential candidate alterations amenable for targeted therapy that may, in the long term, improve the treatment of patients with GBC. The investigation of changes in mRNA expression that reflect differential methylation, the validation of methylation and expression measurements, and the quantification of methylation and gene expression levels after 5-aza-dC treatment of GBC cell lines complement our findings based on patient samples, suggest a functional effect of DNA methylation on the regulation of GBC genes, and add plausibility to the functional role of epigenetic changes in GBC pathogenesis.

Acknowledgment: We are grateful to Angelika Fraas (Institute of Pathology Heidelberg), the Center for Model System and Comparative Pathology (Institute of Pathology Heidelberg), the tissue bank of the National Center for Tumor Diseases (Heidelberg), and Peter Waas (Cancer Epigenomics, German Cancer Research Center, Heidelberg) for excellent technical assistance. Open-access funding was enabled and organized by Projekt DEAL.

Author Contributions: J.L.B. conceived the study design. I.G., A.C., K.M., V.S., E.M., M.T.R., G.T., A.O., B.M., F.G., V.G.C. collected, revised, and analyzed the G.B.C. cohort and clinicopathological data. J.B., S.R., C.B.P., K.M., B.G., M.W., D.S., E.R., S.U.U., T.B.R., O.P. and J.L.B. were involved in experimental design/conduction/interpretation. Data analysis was performed by J.B., F.B., C.B.P., S.U.U. and E.R. J.L.B. and J.B. drafted the manuscript. All authors reviewed the manuscript and approved the final version.

REFERENCES

- Hundal R, Shaffer EA. Gallbladder cancer: epidemiology and outcome. *Clin Epidemiol* 2014;6:99-109.
- Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2018;68:394-424.
- Nemunaitis JM, Brown-Glabeman U, Soares H, Belmonte J, Liem B, Nir I, et al. Gallbladder cancer: review of a rare orphan gastrointestinal cancer with a focus on populations of New Mexico. *BMC Cancer* 2018;18:1-14.
- Roa I, de Aretxabala X. Gallbladder cancer in Chile. *Curr Opin Gastroenterol* 2015;31:269-275.
- Sharma A, Sharma KL, Gupta A, Yadav A, Kumar A. Gallbladder cancer epidemiology, pathogenesis and molecular genetics: recent update. *World J Gastroenterol* 2017;23:3978-3998.
- Lorenzo Bermejo J, Boekstegers F, González Silos R, Marcelain K, Baez Benavides P, Barahona Ponce C, et al. Subtypes of Native American ancestry and leading causes of death: Mapuche ancestry-specific associations with gallbladder cancer risk in Chile. *PLoS Genet* 2017;13:e1006756.
- Stinton LM, Shaffer EA. Epidemiology of gallbladder disease: cholelithiasis and cancer. *Gut Liver* 2012;6:172-187.
- Pilgrim CHC, Groeschl RT, Christians KK, Gamblin TC. Modern perspectives on factors predisposing to the development of gallbladder cancer. *HPB (Oxford)* 2013;15:839-844.
- Espinoza JA, Bizama C, García P, Ferreccio C, Javle M, Miquel JF, et al. The inflammatory inception of gallbladder cancer. *Biochim Biophys Acta* 2016;1865:245-254.
- Sharma P, Bhunia S, Poojary SS, Tekcham DS, Barbhuiya MA, Gupta S, et al. Global methylation profiling to identify epigenetic signature of gallbladder cancer and gallstone disease. *Tumour Biol* 2016;37:14687-14699.
- Takahashi T, Shivapurkar N, Riquelme E, Shigematsu H, Reddy J, Suzuki M, et al. Aberrant promoter hypermethylation of multiple genes in gallbladder carcinoma and chronic cholecystitis. *Clin Cancer Res* 2004;10:6126-6133.
- Letelier P, Brebi P, Tapia O, Roa JC. DNA promoter methylation as a diagnostic and therapeutic biomarker in gallbladder cancer. *Clin Epigenetics* 2012;4:11.
- Li M, Zhang Z, Li X, Ye J, Wu X, Tan Z, et al. Whole-exome and targeted gene sequencing of gallbladder carcinoma identifies recurrent mutations in the ErbB pathway. *Nat Genet* 2014;46:872-876.
- Narayan RR, Creasy JM, Goldman DA, Gonen M, Kandath C, Kundra R, et al. Regional differences in gallbladder cancer pathogenesis: Insights from a multi-institutional comparison of tumor mutations. *Cancer* 2018;125:575-585.
- Weigel C, Veldwijk MR, Oakes CC, Seibold P, Slynko A, Liesenfeld DB, et al. Epigenetic regulation of diacylglycerol kinase alpha promotes radiation-induced fibrosis. *Nat Commun* 2016;7:10893.
- Umu SU, Langseth H, Bucher-Johannessen C, Fromm B, Keller A, Meese E, et al. A comprehensive profile of circulating RNAs in human serum. *RNA Biol* 2018;15:242-250.
- Schubert M, Lindgreen S, Orlando L. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Res Notes* 2016;9:1-7.
- Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012;9:357-359.
- Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* 2015;31:166-169.
- Huang Y, Du Q, Wu W, She F, Chen Y. Rescued expression of WIF-1 in gallbladder cancer inhibits tumor growth and induces tumor cell apoptosis with altered expression of proteins. *Mol Med Rep* 2016;14:2573-2581.
- Lin B, Hong H, Jiang X, Li C, Zhu S, Tang N, et al. WNT inhibitory factor 1 promoter hypermethylation is an early event during gallbladder cancer tumorigenesis that predicts poor survival. *Gene* 2017;622:42-49.
- Roa JC, Anabalón L, Roa I, Melo A, Araya JC, Tapia O, et al. Promoter methylation profile in gallbladder cancer. *J Gastroenterol* 2006;41:269-275.
- Morris TJ, Butcher LM, Feber A, Teschendorff AE, Chakravarthy AR, Wojdacz TK, et al. ChAMP: 450k Chip Analysis Methylation Pipeline. *Bioinformatics* 2014;30:428-430.
- Matsushita S, Onishi H, Nakano K, Nagamatsu I, Imaizumi A, Hattori M, et al. Hedgehog signaling pathway is a potential therapeutic target for gallbladder cancer. *Cancer Sci* 2014;105:272-280.

- 25) **Li J, Wu T**, Lu J, Cao Y, Song N, Yang T, et al. Immunohistochemical evidence of the prognostic value of hedgehog pathway components in primary gallbladder carcinoma. *Surg Today* 2012;42:770-775.
- 26) **Xie F, Xu X**, Xu A, Liu C, Liang F, Xue M, et al. Aberrant activation of Sonic hedgehog signaling in chronic cholecystitis and gallbladder carcinoma. *Hum Pathol* 2014;45:513-521.
- 27) **Lambiv WL, Vassallo I**, Delorenzi M, Shay T, Diserens A-C, Misra A, et al. The Wnt inhibitory factor 1 (WIF1) is targeted in glioblastoma and has a tumor suppressing function potentially by induction of senescence. *Neuro Oncol* 2011;13:736-747.
- 28) Bae S-C, Choi J-K. Tumor suppressor activity of RUNX3. *Oncogene* 2004;23:4336-4340.
- 29) Daskalos A, Logotheti S, Markopoulou S, Xinarianos G, Gosney JR, Kastania AN, et al. Global DNA hypomethylation-induced Δ Np73 transcriptional activation in non-small cell lung cancer. *Cancer Lett* 2010;300:79-86.
- 30) Nakaoka T, Saito Y, Saito H. Aberrant DNA methylation as a biomarker and a therapeutic target of cholangiocarcinoma. *Int J Mol Sci* 2017;18:1111.
- 31) **Nakamura H, Arai Y, Totoki Y, Shirota T, Elzawahry A**, Kato M, et al. Genomic spectra of biliary tract cancer. *Nat Genet* 2015;47:1003-1010.
- 32) **Diaz-Lagares A, Mendez-Gonzalez J**, Hervas D, Saigi M, Pajares MJ, Garcia D, et al. A novel epigenetic signature for early diagnosis in lung cancer. *Clin Cancer Res* 2016;22:3361-3371.
- 33) Liao S-S, Jazag A, Whang EE. HMGA1 is a determinant of cellular invasiveness and in vivo metastatic potential in pancreatic adenocarcinoma. *Cancer Res* 2006;66:11613-11622.
- 34) **Ye L, Li F**, Song Y, Yu D, Xiong Z, Li Y, et al. Overexpression of CDCA7 predicts poor prognosis and induces EZH2-mediated progression of triple-negative breast cancer. *Int J Cancer* 2018;143:2602-2613.
- 35) Omura N, Mizuma M, MacGregor A, Hong S-M, Ayars M, Almaro JA, et al. Overexpression of ankyrin1 promotes pancreatic cancer cell growth. *Oncotarget* 2016;7:34977-34987.
- 36) **Li Q, Lai Q**, He C, Fang Y, Yan Q, Zhang Y, et al. RUNX1 promotes tumour metastasis by activating the Wnt/ β -catenin signaling pathway and EMT in colorectal cancer. *J Exp Clin Cancer Res* 2019;38:334.
- 37) **Kim JJ, Lee SB**, Jang J, Yi S-Y, Kim S-H, Han S-A, et al. WSB1 promotes tumor metastasis by inducing pVHL degradation. *Genes Dev* 2015;29:2244-2257.
- 38) Pikor LA, Lockwood WW, Thu KL, Vucic EA, Chari R, Gazdar AF, et al. YEATS4 is a novel oncogene amplified in non-small cell lung cancer that regulates the p53 pathway. *Cancer Res* 2013;73:7301-7312.
- 39) Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhim R, Getz G. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol* 2011;12:R41.
- 40) Zaidi N, Swinnen JV, Smans K. ATP-citrate lyase: a key player in cancer metabolism. *Cancer Res* 2012;72:3709-3714.
- 41) Albrecht T, Rausch M, Roessler S, Geissler V, Albrecht M, Halske C, et al. HER2 gene (ERBB2) amplification is a low-frequency driver with potential predictive value in gallbladder carcinoma. *Virchows Arch* 2020;476:871-880.
- 42) Mhatre S, Wang Z, Nagrani R, Badwe R, Chiplunkar S, Mittal B, et al. Common genetic variation and risk of gallbladder cancer in India: a case-control genome-wide association study. *Lancet Oncol* 2017;18:535-544.
- 43) **Boekstegers F, Marcelain K**, Barahona Ponce C, Baez Benavides PF, Müller B, de Toro G, et al. ABCB1/4 gallbladder cancer risk variants identified in India also show strong effects in Chileans. *Cancer Epidemiol* 2020;65:101643.
- 44) **Wardell CP, Fujita M**, Yamada T, Simbolo M, Fassan M, Karlic R, et al. Genomic characterization of biliary tract cancers identifies driver genes and predisposing mutations. *J Hepatol* 2018;68:959-969.
- 45) Marchese S, Polo A, Ariano A, Velloso S, Costantini S, Severino L. Aflatoxin B1 and M1: biological properties and their involvement in cancer development. *Toxins (Basel)* 2018;10:214.
- 46) **Nogueira L, Foerster C**, Groopman J, Egner P, Koshiol J, Ferreccio C. Association of aflatoxin with gallbladder cancer in Chile. *JAMA* 2015;313:2075-2077.
- 47) Goepfert B, Konermann C, Schmidt CR, Bogatyrova O, Geiselhart L, Ernst C, et al. Global alterations of DNA methylation in cholangiocarcinoma target the Wnt signaling pathway. *HEPATOLOGY* 2013;59:544-554.
- 48) Harb J, Lin P-J, Hao J. Recent development of Wnt signaling pathway inhibitors for cancer therapeutics. *Curr Oncol Rep* 2019;21:1-9.
- 49) Xie H, Paradise BD, Ma WW, Fernandez-Zapico ME. Recent advances in the clinical targeting of hedgehog/GLI signaling in cancer. *Cells* 2019;8:1-17.
- 50) Capper D, Stichel D, Sahm F, Jones DTW, Schrimpf D, Sill M, et al. Practical implementation of DNA methylation and copy-number-based CNS tumor diagnostics: the Heidelberg experience. *Acta Neuropathol* 2018;136:181-210.
- 51) Stone TJ, Keeley A, Virasami A, Harkness W, Tisdall M, Delgado EI, et al. Comprehensive molecular characterisation of epilepsy-associated glioneuronal tumours. *Acta Neuropathol* 2017;135:115-119.
- 52) **Li M, Liu F, Zhang F, Zhou W, Jiang X**, Yang Y, et al. Genomic ERBB2/ERBB3 mutations promote PD-L1-mediated immune escape in gallbladder cancer: a whole-exome sequencing analysis. *Gut* 2019;68:1024-1033.

Author names in bold designate shared co-first authorship.

Supporting Information

The data set used for the analyses described in this manuscript has been deposited in ArrayExpress <https://www.ebi.ac.uk/arrayexpress/> with accession number E-MTAB-9868. Additional Supporting Information may be found at onlinelibrary.wiley.com/doi/10.1002/hep.31585/supinfo.