DR. JOHANNES BRÄGELMANN (Orcid ID : 0000-0002-1306-2169)
MR. FELIX BOEKSTEGERS (Orcid ID : 0000-0002-0587-7624)
DR. JUSTO LORENZO BERMEJO (Orcid ID : 0000-0002-6568-5333)

Article type : Original

Epigenome-wide analysis of methylation changes in the sequence of gallstone disease, dysplasia, and gallbladder cancer

Johannes Brägelmann[1,2,3], 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*]

[1] Institute of Medical Biometry and Informatics, University of Heidelberg, Heidelberg, Germany
[2] Molecular Pathology, Institute of Pathology & Department of Translational Genomics, University Hospital of Cologne, Cologne, Germany
[3] Mildred Scheel School of Oncology, Cologne, University Hospital Cologne, Medical Faculty, Cologne, Germany
[4]Department of Basic and Clinical Oncology, Medical Faculty, University of Chile, Chile
[5] Institute of Pathology, University Hospital Heidelberg, Heidelberg, Germany
[6] Servicio de Anatomía Patológica, Hospital Clínico de la Universidad de Chile, Santiago, Chile
[7] Servicio de Anatomía Patológica, Hospital Padre Hurtado, Santiago, Chile
[8] Servicio de Anatomía Patológica, Hospital Regional, Talca, Chile

[9] Servicio de Anatomía Patológica, Hospital del Salvador, Santiago, Chile
[10] Servicio de Anatomía Patológica, Hospital de Puerto Montt & Universidad Austral de Chile Sede Puerto Montt, Chile
[11] Servicio de Anatomía Patológica, Hospital Regional, Arica, Chile
[12] Servicio de Oncología Médica, Instituto Nacional del Cáncer, Santiago, Chile
[13] Servicio de Anatomía Patológica, Hospital San Borja Arriarán, Santiago, Chile
[14] Research Unit of Molecular Epidemiology and Institute of Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany
[15] Department of Research, Cancer Registry of Norway, Oslo
[16] Department of Informatics, University of Oslo, Oslo
[17] German Cancer Research Center (DKFZ), Division of Cancer Epigenomics, Heidelberg, Germany.

***Corresponding author**
Justo Lorenzo Bermejo
Statistical Genetics Group, Institute of Medical Biometry and Informatics, University of Heidelberg
Im Neuenheimer Feld 130.3
69120 Heidelberg, Germany
E-mail address: lorenzo@imbi.uni-heidelberg.de. Phone: +496221564180

Keywords:
DNA methylation, epigenetics, gallbladder cancer, Chile

List of abbreviations:
CNV – Copy number variation, FFPE – Formalin-fixed paraffin embedded, GBC – Gallbladder Cancer, GSD – Gallstone disease, MAD – Median absolute deviation

Financial support:
This study was supported by the German Federal Ministry of Education and Research (BMBF, grant 01DN15021); the German Research Foundation (DFG, grant LO 2061/1); the European Union's Horizon 2020 research and innovation programme under grant agreement No 825741; CONICYT-FONDEF IT16I10050; the Biobank of University of Chile, Deutsche Forschungsgemeinschaft (DFG, German Research Foundation, Project-ID 314905040 – SFB/TRR 209) and the Wilhelm Sander-Stiftung (Nr. 2015.111.1) to S.R. J.B. was supported by the Else Kröner-Fresenius Stiftung Memorial Grant (2018_EKMS.35) and the Deutsche Krebshilfe through a Mildred Scheel Nachwuchszentrum Grant (Grant number 70113307). 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.

Conflict of interest statement:
All authors have nothing to disclose.

Acknowledgements:
We are grateful to Angelika Fraas (Institute of Pathology Heidelberg), the Center for Model System and Comparative Pathology (CMCP, Institute of Pathology Heidelberg), the tissue bank of the National Center for Tumor Diseases (NCT, Heidelberg) and Peter Waas (Cancer Epigenomics, DKFZ, Heidelberg) for excellent technical assistance.

Word count: 4893

Abstract

Background & aims: Gallbladder cancer (GBC) is a highly aggressive malignancy of the biliary tract. Most cases of GBC are diagnosed in low- 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: To perform epigenome-wide methylation profiling, genomic DNA extracted from sections of FFPE gallbladder tissue was analyzed using Illumina Infinium MethylationEPIC BeadChips. Pre-processed, 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).

Main results: The number and magnitude of epigenetic changes increased with disease development and predominantly involved the hypermethylation of CpG 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 *CDKN2A*, *MDM2*, *TP53*, and *CCND1*. Gains in the targetable *ERBB2* were detected in 14% of the 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.

Introduction

Gallbladder cancer (GBC; ICD-10 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 worldwide.(1,2) GBC is relatively rare in the USA and Western Europe, but highly prevalent in several countries of South Asia and South America.(1,3) Since most GBC-related deaths occur in low- 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), but also inflammation, toxins, ethnicity and genetic background.(1,5,6) GSD increases GBC risk by 2.4 (gallstones 2.0-2.9 cm in diameter) to 9.2-10.1 (gallstones larger than 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)(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 Infinium MethylationEPIC BeadChips. This 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 the EpiTYPER MassARRAY

technology, investigated the relationship between methylation and gene expression by RNA sequencing (RNAseq) and qRT-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 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 [GOCCHI], and the Biobank of Universidad de Chile. Patients with GSD (cholecystectomy patients without incipient GBC findings), gallbladder dysplasia and GBC were invited to participate. With the exception of two GBC patients with missing gallstone information, all the GBC and dysplasia patients 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, patients with porcelain gallbladder, polyps, non-cholesterol 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 (one GBC patient 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 H&E, 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, EEUU and FFPE QC kits; Illumina®). Fresh-frozen tissue (1-25mg) 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, San Diego, CA, USA) and bisulfite converted (EZ-96 DNA Methylation Kit; Zymo Research, Orange, CA, USA) prior to methylation analysis on an Illumina iScan platform using the Infinium MethylationEPIC BeadChip 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 Supplementary Methods.

EpiTYPER MassARRAY methylation analysis

EpiTYPER MassARRAY (Agena Bioscience, USA) analyses were done as previously described (15) by treating 1.0 µg of DNA using the EZ DNA methylation kit (Zymo Research) followed by PCR-amplification of regions of interest. Primer design: EpiDesigner software (Agena)(**Suppl. Table S1**). Amplicon size was limited to <200bp 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 CpG methyltransferase; Thermo Fisher Scientific, USA) prior to bisulfite conversion.

Gene expression measurements and immunohistochemistry assays

Gene expression was quantified using RNAseq, complemented with qRT-PCR to validate the expression measurements for several candidate genes. RNAseq was performed using NEBNext® Small RNA Library Prep Set for Illumina (Cat. No E7300, New England Biolabs Inc.) with a cut size on the pippin prep (Cat. No CSD3010, Sage Science) covering RNA molecules from 17 to 47 nucleotides, which enables capture of mRNA fragments as shown previously.(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 (log2 TPM). mRNAs with a low count variability (Median absolute deviation <2 counts), and samples with less than 8 million mapped mRNA reads were excluded from the subsequent statistical analyses. The qPCR expression analyses and immunohistochemistry (IHC) assays conducted for the candidate genes are described in the Supplementary Methods.

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 STR profiling. Cells were cultured at 37 °C, 5% CO_2 in RPMI-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 72h. Media were changed daily with fresh 5-aza-dC solution. DNA and RNAseq data for G-415 were generated as described in the **Supplementary Methods.**

Results

Main characteristics of the Chilean cohort of gallbladder disease patients

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 patients), of which 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-value=5.1×10^{-5}; **Figure 1A, Suppl. Table S2**). The difference in median age between GSD and GBC patients was 13.5 years, in accordance with the model of carcinogenesis GSD→dysplasia→GBC.(1) Among GBC patients with available clinicopathological information, only a minority showed advanced T, N, and M stages (25% >T3, 17% >N0, 27% M1), indicating that the majority of investigated GBC patients had localized GBC tumors (**Suppl. Table S2**).

Global methylation differences during GBC development

For 82 out of 89 investigated samples (92%), the hybridization results from Illumina Infinium MethylationEPIC BeadChip fulfilled the quality control filters when we applied the Minfi package. PCA based on markers with the largest methylation variability revealed similar global methylation patterns for GSD and low-grade dysplasia samples (**Figure 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 (**Figure 1C**).

Bonferroni–Holm-corrected p-values from nonparametric 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, Suppl. Table S3**). Interestingly, many of the top 20 markers included CpG sites at or near transcription factors such as *ZNF177* and *TBX15*, including several homeobox genes, for example *BARHL2* and *EN1*. Exemplary box plots with the β methylation values for selected markers are shown in **Suppl. Figure S1A-D**.

In addition to the assessment of significant monotonic changes in methylation, we also used linear regression models with the age and gender of the patients as covariates to calculate group-wise methylation levels (**Table 1, Figure 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 (**Figure 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-value<2×10^{-16})(**Figure 1E**). Markers in gene bodies were under-represented (p-value=2×10^{-267}; **Suppl. Figure S2A**) and those located 1500bp or 200bp within transcription start sites (TSSs; i.e., within promoter regions) were over-represented (p-values: 5×10^{-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-value=5×10^{-63} and TSS200 p-value=8×10^{-142}), and enhancer methylation decreased with GBC progression (**Suppl. Figure S2B**).

To gain a systematic overview, we also performed a 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 (**Suppl. Figure S2C**). Differential methylation was moreover enriched in pathways involved in receptor signaling, including the Wnt and cadherin pathways, which have been previously linked to tissue invasion (cell–cell adhesion) in Chilean GBC (**Suppl. Figure 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 accounting for CpG density and methylation changes (23) to detect 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 *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: *HHIP* (12 CpG sites, mean β log-fold change=0.11, p-value=3.78×10^{-45}) and *PTCH1* (7 CpG sites, mean β log-fold change=0.04, p-value=$5.1\times10^{-}$

[9]). Epigenetic regulation of Hedgehog signaling may thus contribute to GBC carcinogenesis in our cohort.

We also detected hypermethylated promoter DMRs in *WIF1*, *RUNX3*, and *TP73* (**Figure 2A**), which have been classified as tumor suppressor genes and may therefore be of particular mechanistic relevance to GBC.(27-29) **Figure 2B-D** show 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 GBC patients, 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-value=8×10^{-41}, **Figure 2B, top**). Moreover, the methylation in the promoter region and in the body of *WIF1* were negatively correlated (blue area **Figure 2B, middle**), 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 (**Figure 2A**). Its main CpG island comprises 21 markers with monotonically increasing methylation during GBC progression (p-value=1.1×10^{-146}, **Figure 2C, top**). Several clusters of co-regulated 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 **Figure 2C, middle**). This complex pattern of 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-value=2×10^{-56}; **Figure 2D**), which may favor the expression of the long tumor-inhibiting isoform.(29) In addition to monotonic methylation changes during GBC development, we also 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 (**Figure 2E**) and stage (**Suppl. Figure S3A**), and *WIF1* showed higher methylation in metastatic GBC (**Suppl. Figure S3B**).

We also detected DMRs in the promoter regions of other candidate cancer genes (**Figure 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 *RPRM* and *TWIST1*, and *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 (**Suppl. Figure S3C-E**). Hypermethylated promoter regions were also noticed in *DCLK1*, *CDO1*, *ZNF331*, and *ZSCAN18*, frequently found to be hypermethylated in other gastrointestinal tumors, including colorectal cancer.(30) *RPL22*, recently categorized as a potential tumor suppressor in CCA, showed six neighboring hypermethylated markers (**Figure 2A** and **Suppl. Figure S3F**).(31) *CDO1*, *ZNF177*, *BCAT1*, and *TRIM57*, which have been proposed as a four-gene methylation-based diagnostic biomarker for non-small cell lung cancer, also showed hypermethylated promoter DMRs (**Figure 2A**).(32) In addition to the widespread hypermethylation in GBC we also identified genes with promoter hypomethylation and pro-tumorigenic properties reported for other cancer entities. They included HMGA1(33) (mean β log-fold change -0.11, $p=5\times10^{-22}$), ERBB2 (mean β log-fold change -0.035, $p=7\times10^{-4}$), CDCA7(34) (mean β log-fold change -0.11, $p=3\times10^{-27}$), ANK1(35) (mean β log-fold change -0.10, $p=2\times10^{-39}$), RUNX1(36) (mean β log-fold change -0.13, $p=1\times10^{-28}$), WSB1(37) (mean β log-fold change -0.105, $p=2\times10^{-8}$) and YEATS4(38) (mean β log-fold change -0.07, $p=5\times10^{-8}$). Our analyses thereby detected novel genes with a potential tumor promoting function in GBC, which constitute bona fide candidates for experimental follow-up studies. All 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 81 of 89 samples (91%) passing QC filters leading to highly

correlated p-values (r=0.98) and estimated methylation differences (r=0.80), and in largely similar DMR-Volcano plots (**Suppl. Figure S4A-C**).

Interestingly, we identified methylation changes according to tumor stage and grade in several candidate genes (**Figure 2E**, **Suppl. Figure S3**), but the comparison of G1 versus G2+G3 GBC tumors, high-grade dysplasia versus GBC and the investigation of the sequence high-grade dysplasia→G1→G2→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 analysis (**Figure 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→low-grade dysplasia→high-grade dysplasia→GBC the frequency of altered CN segments per sample increased (p-value = 6.8×10^{-8}, **Figure 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 (**Figure 3B-F**). Among the genes most frequently affected were the tumor suppressors *CDKN2A* and *TP53*, with genomic losses in 8 (29%) and 5 (18%) GBC samples, respectively (**Figure 3B, D**). We found frequent copy number gains of *MDM2* (GBC: 21%; high-grade dysplasia 22%), a major negative regulator of *TP53* and of *CCND1*, which drives cell cycle progression (**Figure 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 above. 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 GISTIC2(39) revealed recurrent alterations in *MDM2*, *YEATS4*, *CCND1*, and *CDKN2A* (**Suppl. Figure 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 (*GAST*), eukaryotic translation initiation factor 1 (*EIF1*), and ATP citrate lyase (*ACLY*), 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 novel 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), *CDK4* (gains in 22% of high-grade dysplasia samples), *FGFR3*, *ERBB3*, *MET*, and *CDK6* (**Figure 3B**). The fibroblast growth factor ligands *FGF3*, *FGF4*, and *FGF19* were co-amplified in the four (14%) GBC samples with *CCND1* gains (**Figure 3B, Suppl. Figure S5**), suggesting that tailored combination therapy could be indicated in these GBC patients. 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.

Validation of methylation measurements, and investigation of the relationship between methylation and mRNA expression

To validate the identified methylation differences using Illumina Infinium MethylationEPIC BeadChip, 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 and GBC n=10). The side-by-side plots in **Figure 4A** showed good overall consistency between the results based on epigenome-wide EPIC and candidate-marker MassARRAY measurements (n=19 matched samples; **Suppl. Table S4**).

We next applied small RNAseq, which allowed us to quantify the mRNA expression of around 12,000 genes per sample on average (**Suppl. Figure S6A**), complemented with RT-qPCR to validate the expression differences between GSD and GBC samples for the candidate genes *CDO1*, *RUNX3* and *HMGA1*. The side-by-side plots in **Figure 4B** revealed a good overall consistency between RNAseq and qPCR results. We noticed a negative correlation between

RNAseq expression and EPIC methylation for *CDO1*, *HMGA1* (cg08335767) and *RUNX3* in GBC patients (n=20; **Figure 4C**) as well as in the complete set of investigated GSD, dysplasia and GBC samples (n=51; **Suppl. Figure S6B**). In agreement with RNAseq and qPCR results, IHC analysis of the available slides (n=11) showed a higher CDO1 expression in GSD than GBC patients (p=0.01; **Suppl. Table S5**) and a negative relationship between CDO1 expression and methylation (p=0.02, **Suppl. Figure S7A** and **S7B**). However, IHC for CDO1 in an independent group of ten Chilean GBC patients with paired tumor and adjacent non-tumor samples did not show differences in CDO1 expression (increased expression in n=3, equal expression in n=4, and decreased expression in n=3 pairs, **Suppl. Figure 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 (**Figure 5A**). Accordingly, the three genes *HAND2*, *FEZF2* and *EDNRB*, hypermethylated in our DMR analyses (**Table 2**) and with available RNAseq mRNA expression data, showed decreasing expression levels towards GBC (**Figure 5B**). We also observed decreasing expression values for the hypermethylated genes *APC* (Wnt signaling), *HHIP* (Hedgehog signaling) and *ZSCAN18* (**Figure 5C**), while the hypomethylated genes *HMGA1* and *ERBB2* showed increasing expression levels (**Figure 4B** and **Figure 5C**, respectively). Moreover, several genes with gradual methylation changes along the progression from GSD to GBC according to JT-tests (**Table 1**) showed the contrary tendencies in mRNA expression (**Suppl. Table S6**), further indicating a functional relevance of the observed methylation changes. The previously reported hypermethylation and decreased expression of *FBN1*, *SOD3* and *LPP* in Indian GBC patients was also apparent in our cohort of Chilean GBC patients (10).(**Suppl. Figure 6C**).

In accordance with the EPIC array-based CNV data, *MDM2* and *ERBB2* were overexpressed in samples that showed CN gains (**Figure 5D**). The validation of the detected *ERBB2* gains in two patients using *ERBB2* dc-CISH lends additional support to our results (**Figure 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* (**Figure 6A**). By contrast, hypomethylated genes in GBC with reported pro-tumorigenic effects (e.g. *YEATS4*, *WSB1*, *CDCA7*, *ANK1* and *HMGA1*) showed in general low promoter methylation and higher expression in G-415.

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 **Figure 6C** depicts the expression levels of *CDO1*, *RUNX3*, *TP73* and *HMGA1* after 72h treatment. Compared to controls, 5-aza-dC treatment reduced the methylation of *CDO1*, *RUNX3* and *TP73* in both cell lines, while the methylation of *HMGA1* (hypomethylated in GBC) was unaltered (**Figure 6B**). Accordingly, 5-aza-dC treatment resulted in an increased expression of *CDO1* and *RUNX3* in both cell lines, supporting a transcriptional de-repression by decreased DNA methylation (**Figure 6C**). The lower methylation of *TP73* after treatment translated into an increased expression in G-415, but no expression change was noticed in OZ. As expected, the changes of HMGA1 expression were not statistically significant (**Suppl. Table S4).**

Discussion

In the present study we report for the first time on epigenome-wide changes in methylation and CNVs along the sequence GSD→low-grade dysplasia→high-grade dysplasia→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 GBC patients,(42) and applied a different study design (fresh-frozen tumor, adjacent non-tumor and gallstone samples). (10,43) An enhanced understanding of the molecular alterations occurring during GBC

development in regions with a high incidence of both gallstone disease 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 well established GBC risk factors in Amerindian and European populations is GSD, which is considered to induce GBC via inflammation following a dysplasia→carcinoma sequence.(1,5,6,9,12) This holds true especially in Chile, and was reflected into 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→dysplasia→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 GSD and GBC patients is in accordance with the time postulated for transformation of GSD to GBC.(1,9,11,12) Similarly, unsupervised PCA analysis clustered GBC samples farthest 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*, *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 GBC patients than in healthy community controls, suggesting greater aflatoxin exposure in GBC patients.(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 seems 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 an increased methylation (quantified both by EPIC and MassArray) and simultaneously an increased expression (according to RNAseq and qPCR) 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 increasing 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 novel GBC treatment strategies.

In addition to epigenetic information, methylation arrays offer the opportunity to investigate copy number changes.(50,51) Interestingly, we not only observed frequent loss of *TP53*, but also co-occuring 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 CN 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 new 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, to our knowledge this is the first study to assess 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 novel 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 GBC patients. The investigation of changes in mRNA expression that reflect differential methylation, the conducted 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.

References

1. Hundal R, Shaffer EA. Gallbladder cancer: epidemiology and outcome. Clin Epidemiol. 2014;6:99–109.
2. 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.
3. **Nemunaitis JM, Brown-Glabeman U**, Soares H, Belmonte J, Ben Liem, Nir I, et al. Gallbladder cancer: review of a rare orphan gastrointestinal cancer with a focus on populations of New Mexico. 2018;18:1–14.
4. Roa I, de Aretxabala X. Gallbladder cancer in Chile. Current Opinion in Gastroenterology. 2015;31:269–275.
5. 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.
6. **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.
7. Stinton LM, Shaffer EA. Epidemiology of Gallbladder Disease: Cholelithiasis and Cancer. Gut Liver. 2012;6:172–187.
8. Pilgrim CHC, Groeschl RT, Christians KK, Gamblin TC. Modern perspectives on factors predisposing to the development of gallbladder cancer. HPB Journal. 2013;15:839–844.
9. Espinoza JA, Bizama C, García P, Ferreccio C, Javle M, Miquel JF, et al. The inflammatory inception of gallbladder cancer. BBA - Reviews on Cancer. 2016;1865:245–254.
10. **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. Tumor Biology. 2016;37:14687–14699.
11. 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.
12. Letelier P, Brebi P, Tapia O, Roa JC. DNA promoter methylation as a diagnostic and therapeutic biomarker in gallbladder cancer. Clin Epigenet. 2012;4:11.
13. **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.
14. Narayan RR, Creasy JM, Goldman DA, Gonen M, Kandoth C, Kundra R, et al. Regional differences in gallbladder cancer pathogenesis: Insights from a multi-institutional comparison of tumor mutations. Cancer [Internet]. 2018;125:575–585.
15. Weigel C, Veldwijk MR, Oakes CC, Seibold P, Slynko A, Liesenfeld DB, et al. Epigenetic regulation of diacylglycerol kinase alpha promotes radiation-induced fibrosis. Nature Communications. 2016;7:10893.
16. Umu SU, Langseth H, Bucher-Johannessen C, Fromm B, Keller A, Meese E, et al. A comprehensive profile of circulating RNAs in human serum. 2018:1–9.
17. Schubert M, Lindgreen S, Orlando L. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. BMC Res Notes. 2016:1–7.
18. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9:357–359.
19. Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics. 2015;31:166–169.
20. 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 Report. 2016;14:2573–2581.
21. 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.
22. 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.

23. 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.
24. 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 dNp73 transcriptional activation in non-small cell lung cancer. 2010;300:79–86.
30. Nakaoka T, Saito Y, Saito H. Aberrant DNA Methylation as a Biomarker and a Therapeutic Target of Cholangiocarcinoma. IJMS. 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. Clinical Cancer Research. 2016;22:3361–3371.
33. Liau S-S, Jazag A, Whang EE. HMGA1 is a determinant of cellular invasiveness and in vivo metastatic potential in pancreatic adenocarcinoma. Cancer Research. 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, Almario 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 signalling 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 Research. 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 Research. 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. 2019;1–10.
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. Journal of Hepatology. 2018;68:959–969.
45. Marchese S, Polo A, Ariano A, Velotto S, Costantini S, Severino L. Aflatoxin B1 and M1: Biological Properties and Their Involvement in Cancer Development. Toxins.

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: The Journal of the American Medical Association. 2015;313:2075–2077.
47. Goeppert 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. 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 methylationand copy-number-based CNS tumor diagnostics: the Heidelberg experience. Acta Neuropathologica. 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 Neuropathologica. 2017;135:115–129.
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.

Figure legends

Figure 1: Global methylation differences during GBC development

A) Age distribution in the four investigated groups of patients. **B)** Principal component analysis (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 –log10 Bonferroni–Holm-corrected JT p-values. **E)** Distribution of functional genetic elements for differentially (left) and nondifferentially (right) methylated markers (p-value from Fisher exact test). **F)** Distribution of methylation differences between GSD and GBC samples of significant markers according to their location in gene bodies or 1500 bp or 200 bp upstream of the nearest transcription start site (TSS; p-values from two sample U-tests).

Figure 2: Differential methylation of candidate genes

A) Candidate cancer genes with differentially methylated regions (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 ANOVA tests adjusting for baseline CpG differences; green = GSD, light green = LG-dysplasia, light red = HG-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)

Figure 3: Copy number profiles in the investigated Chilean cohort

A) Number of significant copy number (CN) segments in the four investigated groups of patients. **B)** CN variations (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. **C–F)** Typical genome-wide CN 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.

Figure 4: Validation of methylation differences and relationship between DNA methylation and RNA expression

A) Box plots with β methylation values in samples from GSD (n=10) and GBC (n=10) patients for selected CpGs. The 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 vs. GBC). **B)** Gene expression levels determined by RNAseq (log2 transcripts per million; TPM; left y-axis) and RT-qPCR (Ct values normalized to two housekeeping genes to quantify the relative expression in arbitrary units; right y-axis) for GSD (n=8 to10) and GBC (n=9 to 10) validation samples. P-values are based on two-sample U-tests. **C)** Correlation between β values for the CpGs shown in **Figure 4A** 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.

Figure 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 **Figure 5D**. Right picture is from a second patient with a predicted ERBB2 gain. Scale bar 20µm.

Figure 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 (TPM) by RNA-sequencing 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 **Panel A** for the GBC cell lines G-415 and OZ after 72h 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 RT-qPCR for the GBC cell lines G-415 and OZ after 72h treatment with 5-aza-dC or DMSO as negative control (n=3). Ct values were normalized to three housekeeping genes and fold-changes were calculated using the 2^{-ddCt} method. Displayed are mean+SEM. The p-values in the upper part of the panel are based on t-tests.

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	p-value[1]	Mean β-value[2] in GSD samples	Mean β-difference[2] LGD vs. GSD	Mean β-difference[2] HGD vs. GSD	Mean β-difference[2] 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	WWOX	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	WNK4	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)

[1] Bonferroni–Holm-adjusted p-value from Jonckheere–Terpstra test

[2] 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 first gene is shown.

GSD = gallstone disease; LGD = low-grade dysplasia; HGD = high-grade dysplasia; GBC = gallbladder cancer

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 p-value
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

Figure 1

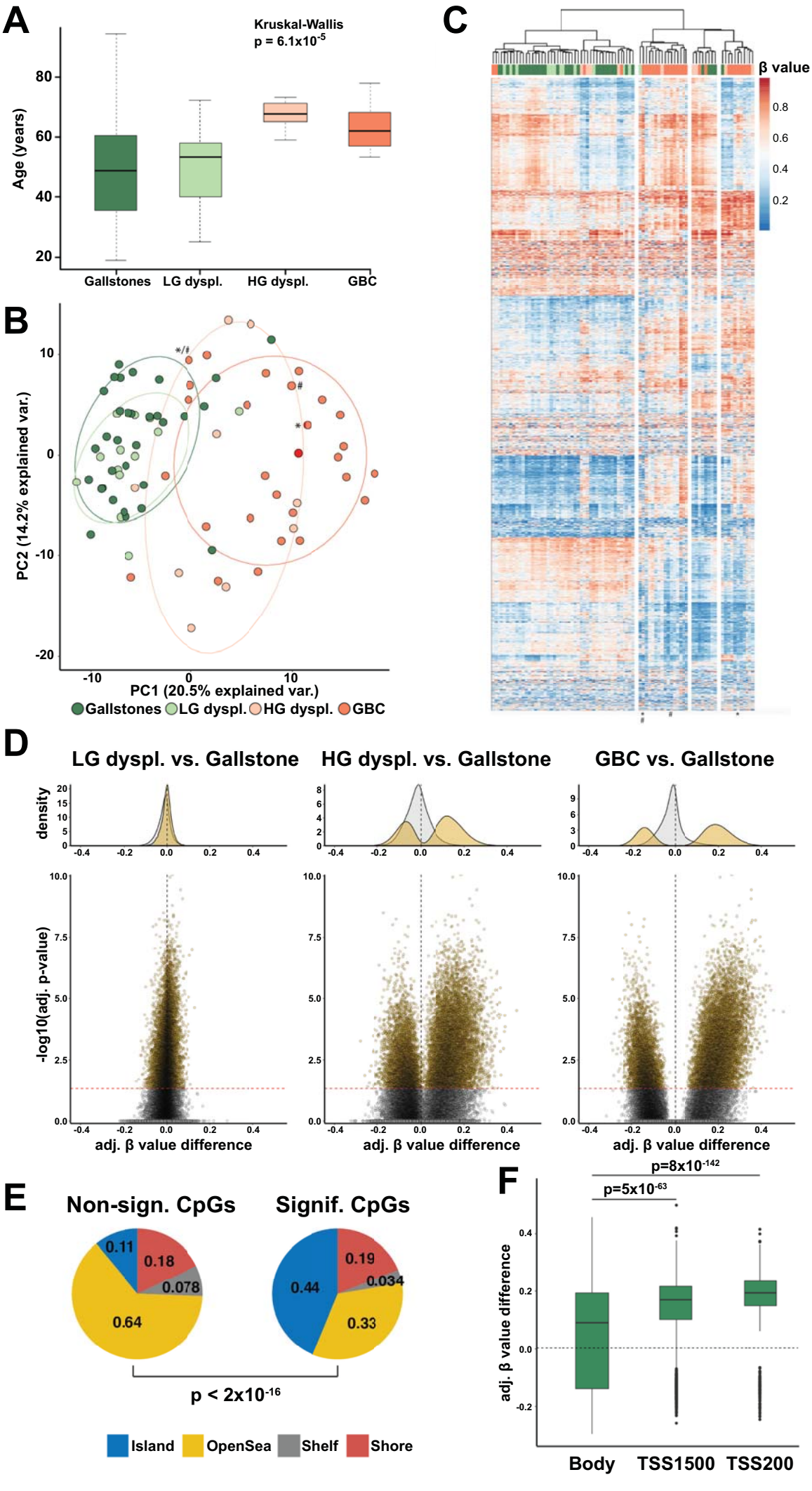

A
Kruskal-Wallis
p = 6.1x10-5
Age (years)
Gallstones
LG dyspl.
HG dyspl.
GBC
B
PC2 (14.2% explained var.)
PC1 (20.5% explained var.)
Gallstones
LG dyspl.
HG dyspl.
GBC
C
β value
D
LG dyspl. vs. Gallstone
HG dyspl. vs. Gallstone
GBC vs. Gallstone
density
-log10(adj. p-value)
adj. β value difference
E
Non-sign. CpGs
Signif. CpGs
0.11
0.18
0.078
0.64
0.19
0.034
0.44
0.33
p < 2x10-16
Island
OpenSea
Shelf
Shore
F
p=8x10-142
p=5x10-63
adj. β value difference
Body
TSS1500
TSS200

Figure 2

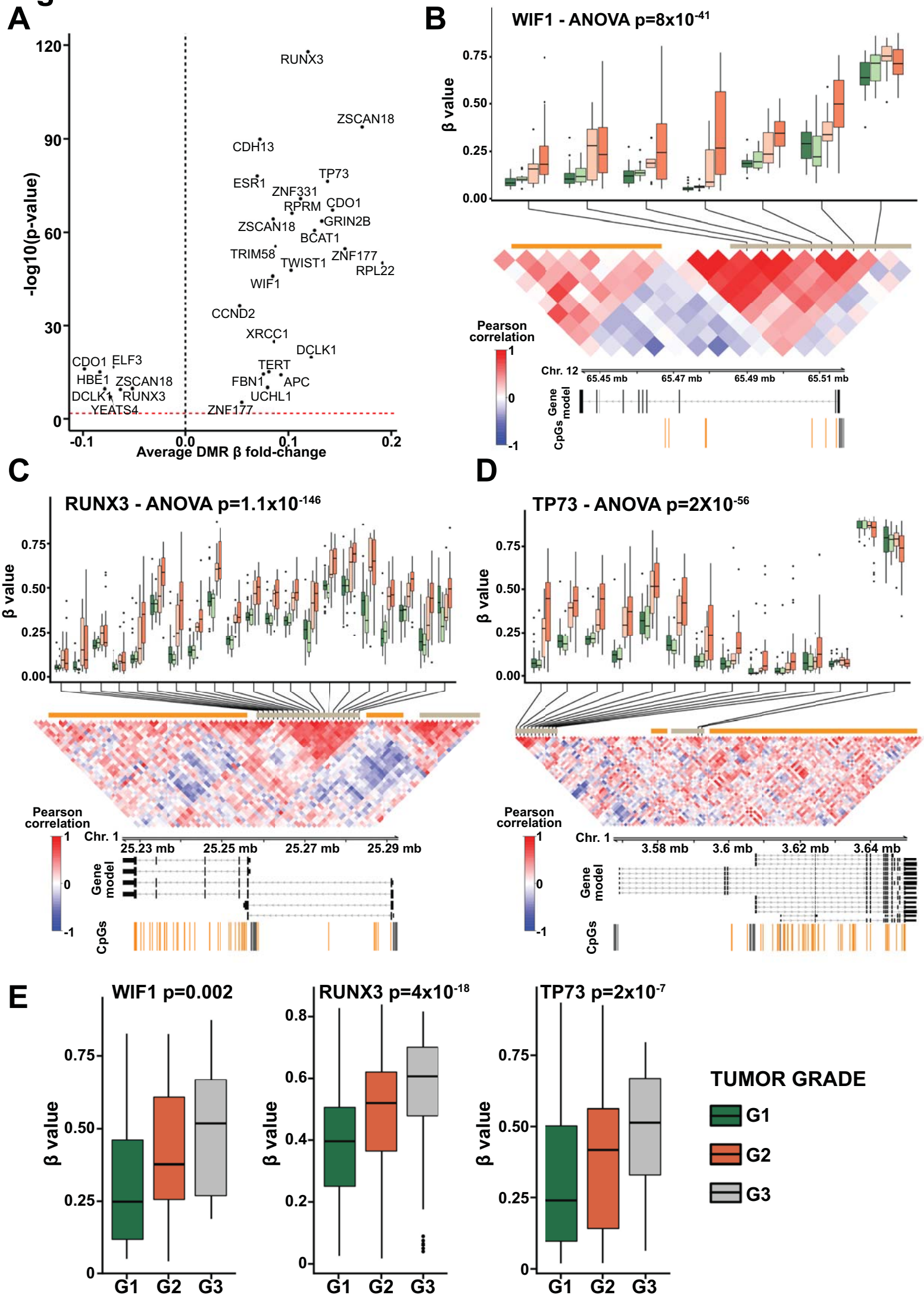

A
-log10(p-value)
Average DMR β fold-change
RUNX3
ZSCAN18
CDH13
TP73
ESR1
ZNF331
RPRM
CDO1
ZSCAN18
GRIN2B
BCAT1
TRIM58
ZNF177
TWIST1
RPL22
WIF1
CCND2
XRCC1
DCLK1
CDO1
ELF3
TERT
HBE1
ZSCAN18
FBN1
APC
DCLK1
RUNX3
UCHL1
YEATS4
ZNF177
B
WIF1 - ANOVA p=8x10-41
β value
Pearson correlation
Chr. 12
65.45 mb
65.47 mb
65.49 mb
65.51 mb
Gene model
CpGs
C
RUNX3 - ANOVA p=1.1x10-146
β value
Pearson correlation
Chr. 1
25.23 mb
25.25 mb
25.27 mb
25.29 mb
Gene model
CpGs
D
TP73 - ANOVA p=2X10-56
β value
Pearson correlation
Chr. 1
3.58 mb
3.6 mb
3.62 mb
3.64 mb
Gene model
CpGs
E
WIF1 p=0.002
RUNX3 p=4x10-18
TP73 p=2x10-7
β value
G1
G2
G3
TUMOR GRADE
G1
G2
G3

Figure 3

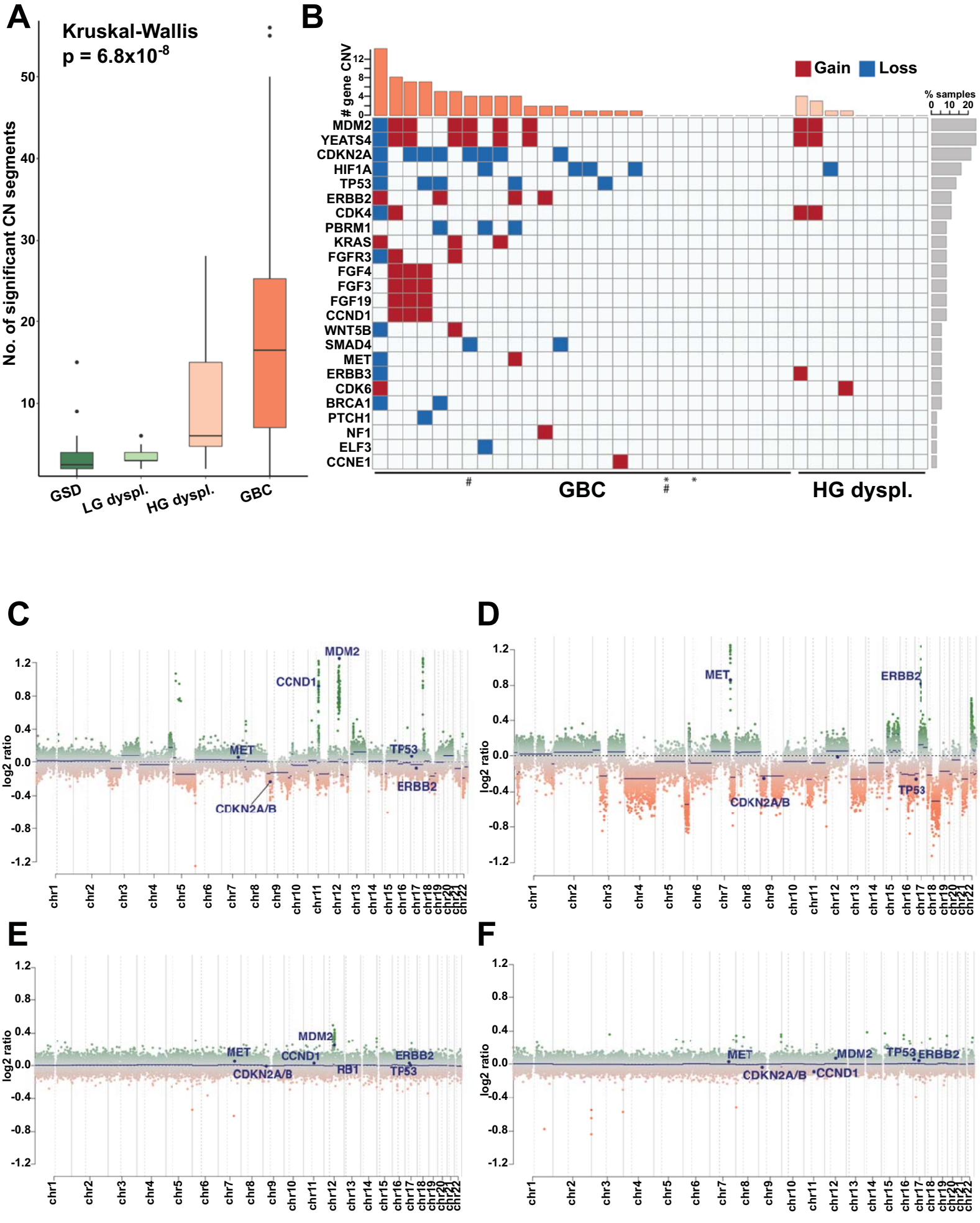
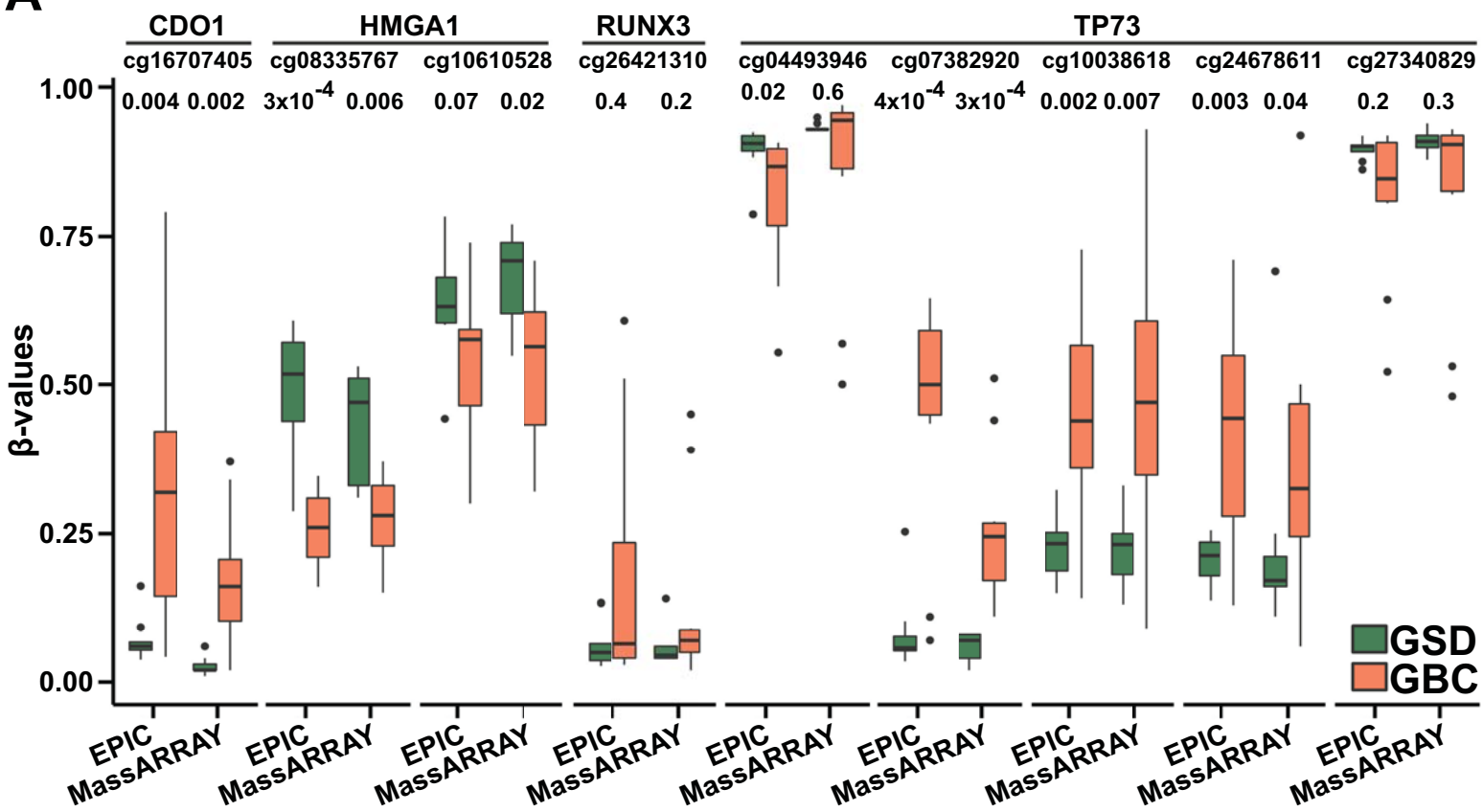
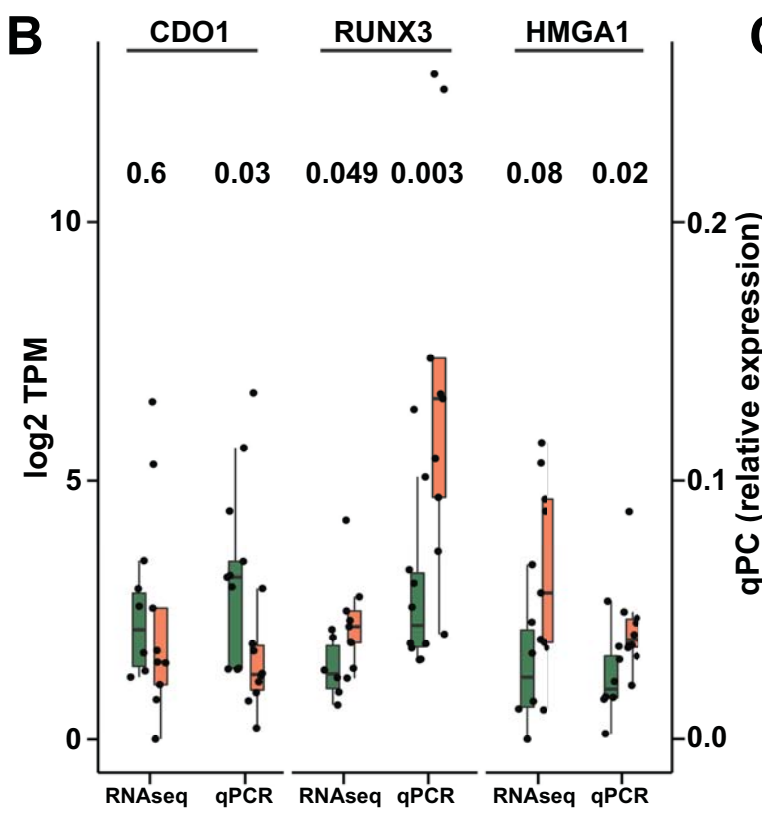

A
Kruskal-Wallis
p = 6.8x10-8
No. of significant CN segments
GSD
LG dyspl.
HG dyspl.
GBC
B
gene CNV
Gain
Loss
% samples
MDM2
YEATS4
CDKN2A
HIF1A
TP53
ERBB2
CDK4
PBRM1
KRAS
FGFR3
FGF4
FGF3
FGF19
CCND1
WNT5B
SMAD4
MET
ERBB3
CDK6
BRCA1
PTCH1
NF1
ELF3
CCNE1
GBC
HG dyspl.
C
log2 ratio
MDM2
CCND1
MET
TP53
ERBB2
CDKN2A/B
D
MET
ERBB2
TP53
CDKN2A/B
E
MDM2
MET
CCND1
ERBB2
CDKN2A/B
RB1
TP53
F
MET
MDM2
TP53
ERBB2
CDKN2A/B
CCND1

Figure 4


A
CDO1
HMGA1
RUNX3
TP73
cg16707405
cg08335767
cg10610528
cg26421310
cg04493946
cg07382920
cg10038618
cg24678611
cg27340829
0.004 0.002
3x10-4 0.006
0.07 0.02
0.4 0.2
0.02 0.6
4x10-4 3x10-4
0.002 0.007
0.003 0.04
0.2 0.3
β-values
1.00
0.75
0.50
0.25
0.00
GSD
GBC
EPIC
MassARRAY


B
CDO1
RUNX3
HMGA1
0.6 0.03
0.049 0.003
0.08 0.02
log2 TPM
qPC (relative expression)
RNAseq
qPCR

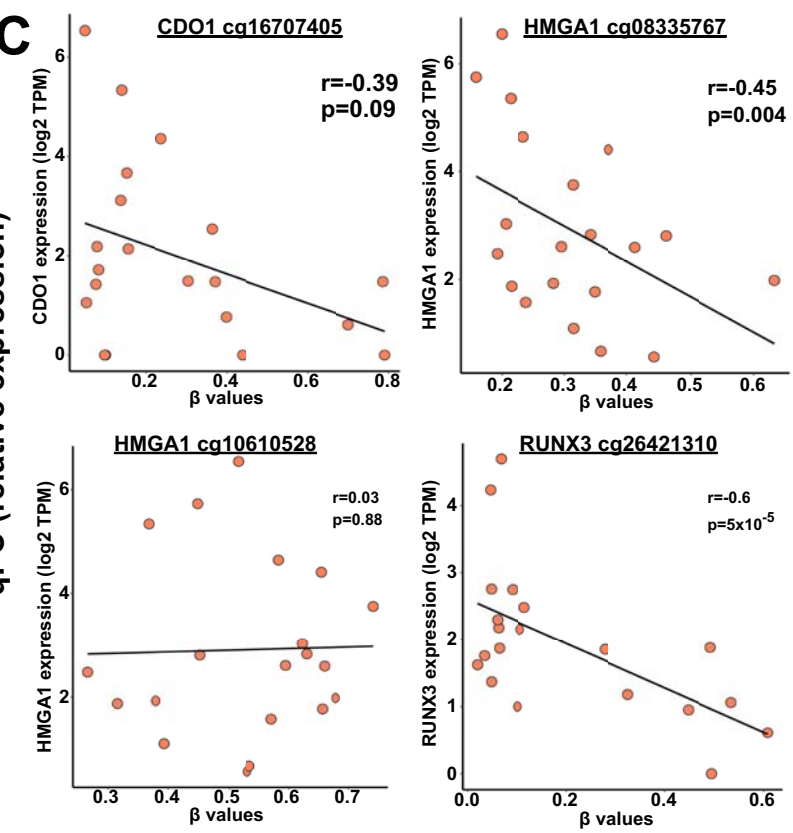

C
CDO1 cg16707405
r=-0.39
p=0.09
HMGA1 cg08335767
r=-0.45
p=0.004
HMGA1 cg10610528
r=0.03
p=0.88
RUNX3 cg26421310
r=-0.6
p=5x10-5
β values

Figure 5

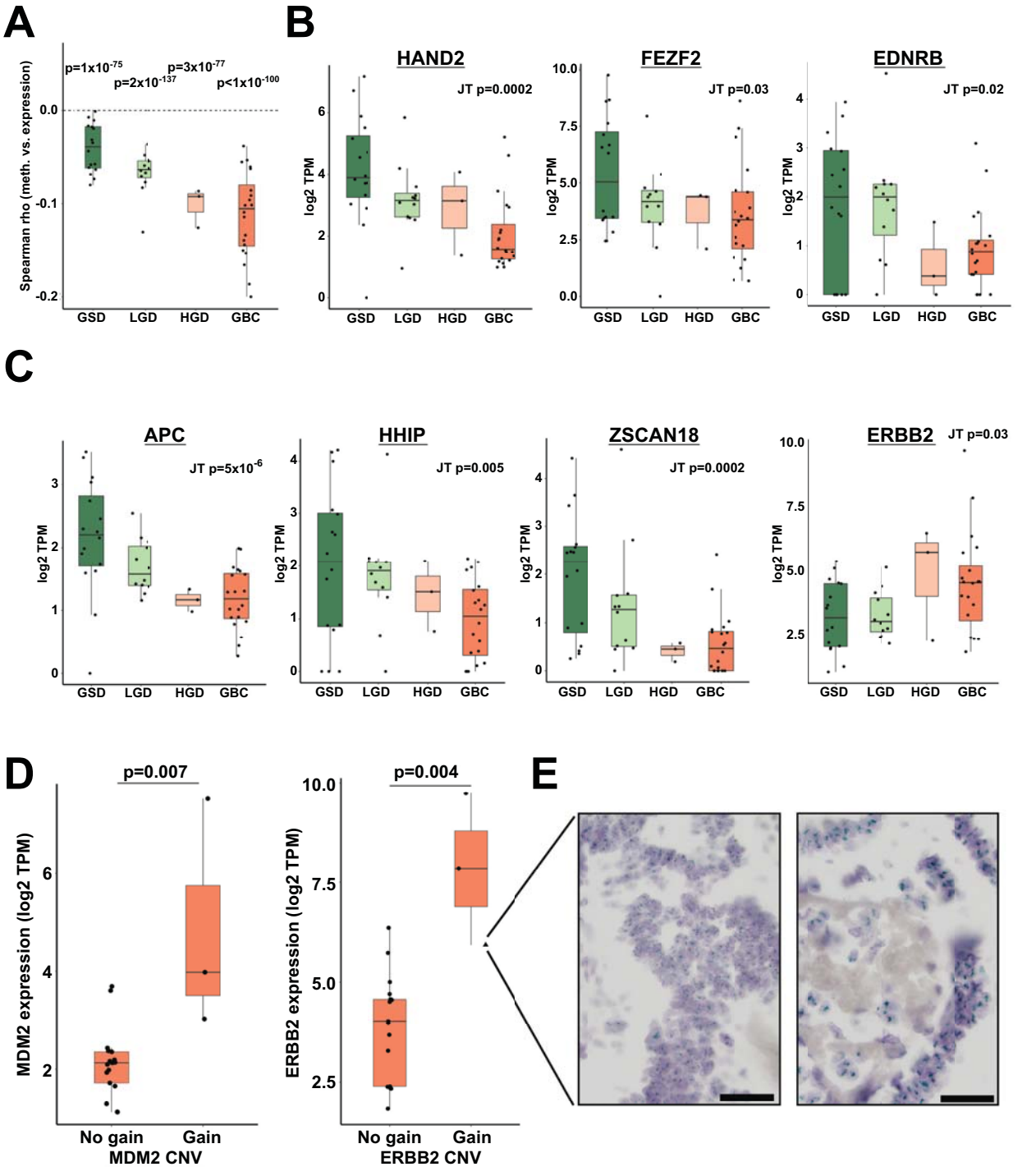
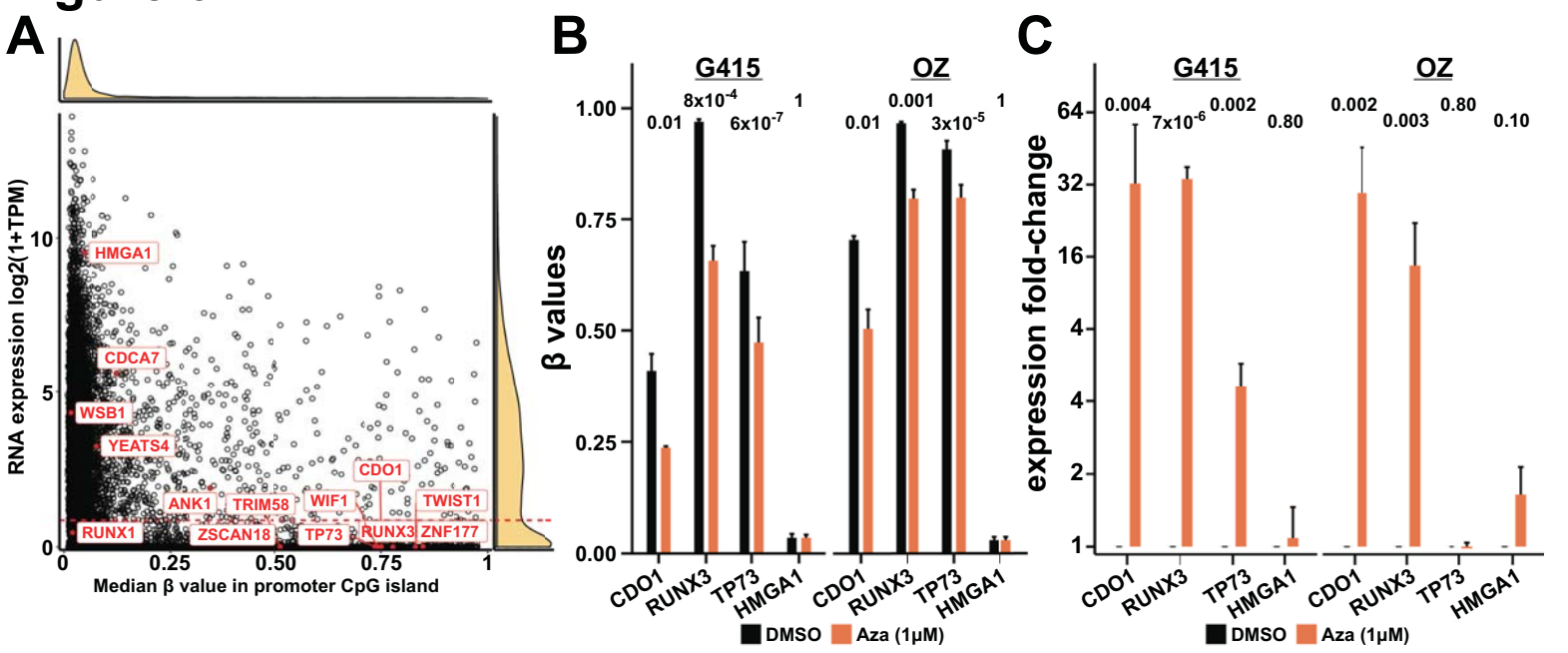

A
B
C
D
E
p=1x10-75
p=2x10-137
p=3x10-77
p<1x10-100
Spearman rho (meth. vs. expression)
GSD
LGD
HGD
GBC
HAND2
JT p=0.0002
FEZF2
JT p=0.03
EDNRB
JT p=0.02
log2 TPM
APC
JT p=5x10-6
HHIP
JT p=0.005
ZSCAN18
JT p=0.0002
ERBB2
JT p=0.03
p=0.007
MDM2 expression (log2 TPM)
No gain
Gain
MDM2 CNV
p=0.004
ERBB2 expression (log2 TPM)
ERBB2 CNV

Figure 6


A
RNA expression log2(1+TPM)
Median β value in promoter CpG island
HMGA1
CDCA7
WSB1
YEATS4
CDO1
ANK1
TRIM58
WIF1
TWIST1
RUNX1
ZSCAN18
TP73
RUNX3
ZNF177
B
G415
OZ
β values
8x10-4
6x10-7
0.001
3x10-5
0.01
CDO1
RUNX3
TP73
HMGA1
DMSO
Aza (1μM)
C
G415
OZ
expression fold-change
0.004
7x10-6
0.002
0.80
0.003
0.10
CDO1
RUNX3
TP73
HMGA1
DMSO
Aza (1μM)