

RESEARCH

Open Access



# Highly recurrent CBS epimutations in gastric cancer CpG island methylator phenotypes and inflammation

Nisha Padmanabhan<sup>1</sup>, Huang Kie Kyon<sup>1†</sup>, Arnoud Boot<sup>2†</sup>, Kevin Lim<sup>1</sup>, Supriya Srivastava<sup>3</sup>, Shuwen Chen<sup>4</sup>, Zhiyuan Wu<sup>5</sup>, Hyung-O K Lee<sup>6</sup>, Vineeth T. Mukundan<sup>7</sup>, Charlene Chan<sup>7</sup>, Yarn Kit Chan<sup>1</sup>, Ong Xuewen<sup>1</sup>, Jason J. Pitt<sup>7</sup>, Zul Fazreen Adam Isa<sup>1</sup>, Manjie Xing<sup>1</sup>, Ming Hui Lee<sup>1</sup>, Angie Lay Keng Tan<sup>1</sup>, Shamaine Ho Wei Ting<sup>1</sup>, Micah A. Luftig<sup>8</sup>, Dennis Kappel<sup>7,9</sup>, Warren D. Kruger<sup>6</sup>, Jinsong Bian<sup>5,10</sup>, Ying Swan Ho<sup>4</sup>, Ming Teh<sup>11</sup>, Steve George Rozen<sup>2</sup> and Patrick Tan<sup>1,7,12,13,14,15\*</sup> 

\* Correspondence: [gmstan@duke-nus.edu.sg](mailto:gmstan@duke-nus.edu.sg)

<sup>†</sup>Huang Kie Kyon and Arnoud Boot contributed equally to this work.

<sup>1</sup>Programme in Cancer and Stem Cell Biology, Duke-NUS Medical School, 8, College road, Singapore 169857, Singapore

<sup>7</sup>Cancer Science Institute of Singapore, National University of Singapore, Singapore 117599, Singapore

Full list of author information is available at the end of the article

## Abstract

**Background:** CIMP (CpG island methylator phenotype) is an epigenetic molecular subtype, observed in multiple malignancies and associated with the epigenetic silencing of tumor suppressors. Currently, for most cancers including gastric cancer (GC), mechanisms underlying CIMP remain poorly understood. We sought to discover molecular contributors to CIMP in GC, by performing global DNA methylation, gene expression, and proteomics profiling across 14 gastric cell lines, followed by similar integrative analysis in 50 GC cell lines and 467 primary GCs.

**Results:** We identify the cystathionine beta-synthase enzyme (CBS) as a highly recurrent target of epigenetic silencing in CIMP GC. Likewise, we show that CBS epimutations are significantly associated with CIMP in various other cancers, occurring even in premalignant gastroesophageal conditions and longitudinally linked to clinical persistence. Of note, CRISPR deletion of CBS in normal gastric epithelial cells induces widespread DNA methylation changes that overlap with primary GC CIMP patterns. Reflecting its metabolic role as a gatekeeper interlinking the methionine and homocysteine cycles, CBS loss in vitro also causes reductions in the anti-inflammatory gasotransmitter hydrogen sulfide (H<sub>2</sub>S), with concomitant increase in NF-κB activity. In a murine genetic model of CBS deficiency, preliminary data indicate upregulated immune-mediated transcriptional signatures in the stomach.

**Conclusions:** Our results implicate CBS as a bi-faceted modifier of aberrant DNA methylation and inflammation in GC and highlights H<sub>2</sub>S donors as a potential new therapy for CBS-silenced lesions.

**Keywords:** Gastric cancer, CBS, CIMP, Inflammation



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

## Background

The CpG island methylator phenotype (CIMP) is a molecular subtype observed in several malignancies, characterized by aberrant, pervasive, and genome-wide DNA hypermethylation of CpG islands [CGIs]. CGI hypermethylation at gene promoter regions can result in transcriptional gene silencing (“epimutations”), and examples of cancer driver genes inactivated by epimutation include *MLH1* [mutL homolog 1] which can cause microsatellite instability [MSI] and *CDKN2A* [cyclin-dependent kinase inhibitor 2A] resulting in defective cell cycle regulation [1–3]. Recent pan-cancer studies comparing CIMP across multiple tumor types have revealed striking tissue-specific methylation patterns, with minimal conserved hypermethylation [2]. At the pathway level however, analysis of CIMP in various cancers has highlighted common denominators such as the enrichment of PRC2 [polycomb repressive complex 2] targets [4, 5].

Clinically, CIMP in different tumor types has been associated with age, gender, tumor location, histological subtype, survival, and exposure to smoking and infectious pathogens [6]. However, molecular mechanisms driving CIMP remain obscure for most cancer types with very few exceptions. Gain-of-function *IDH1* [isocitrate dehydrogenase (NADP(+)) 1] and loss-of-function *TET* [ten–eleven translocation] mutations have been shown to induce CIMP in one more than one cancer type [2]. Other genes and pathways implicated in CIMP include *SDHB* [succinate dehydrogenase complex iron sulfur subunit B] in paraganglioma [7], *KRAS* [KRAS proto-oncogene, GTPase], and activating *BRAF* [B-Raf proto-oncogene, serine/threonine kinase] mutations in colorectal cancer [8, 9].

Gastric cancer [GC] is a leading cause of global cancer mortality and known to display two distinct CIMP groups—“gastric CIMP” associated with MSI, and “EBV [Epstein-Barr Virus] CIMP” associated with EBV positivity [10]. To date, (epi)genetic drivers of GC CIMP remain largely unknown, although EBV infection is known to induce hypermethylation in vitro [11–13]. Despite their distinct etiologies, gastric CIMP and EBV-CIMP have been reported to share common DNA hypermethylation patterns [11]. However, whether these common hypermethylation patterns are functionally required to contribute and/or sustain features of CIMP, or are mere bystander phenotypes, remains poorly understood. Here, through genome-wide studies of DNA methylation and gene expression patterns in GC cell lines and patient cohorts, we discovered *CBS* [cystathionine beta-synthase] promoter hypermethylation as a highly recurrent epimutation in GC CIMP. We show that loss of *CBS* activity in vitro recapitulates a subset of methylation in GC CIMP tumors, and that besides GC, epimutations at *CBS* are also associated with CIMP in other tumor types, implying a broader role in cancer. Surprisingly, *CBS* loss is also associated with an inflammatory response that may be linked in part to *CBS*’s ability to regulate H<sub>2</sub>S production.

## Results

### Integrative methylome, transcriptome, and proteome analysis identify highly recurrent *CBS* epimutations in GC CIMP

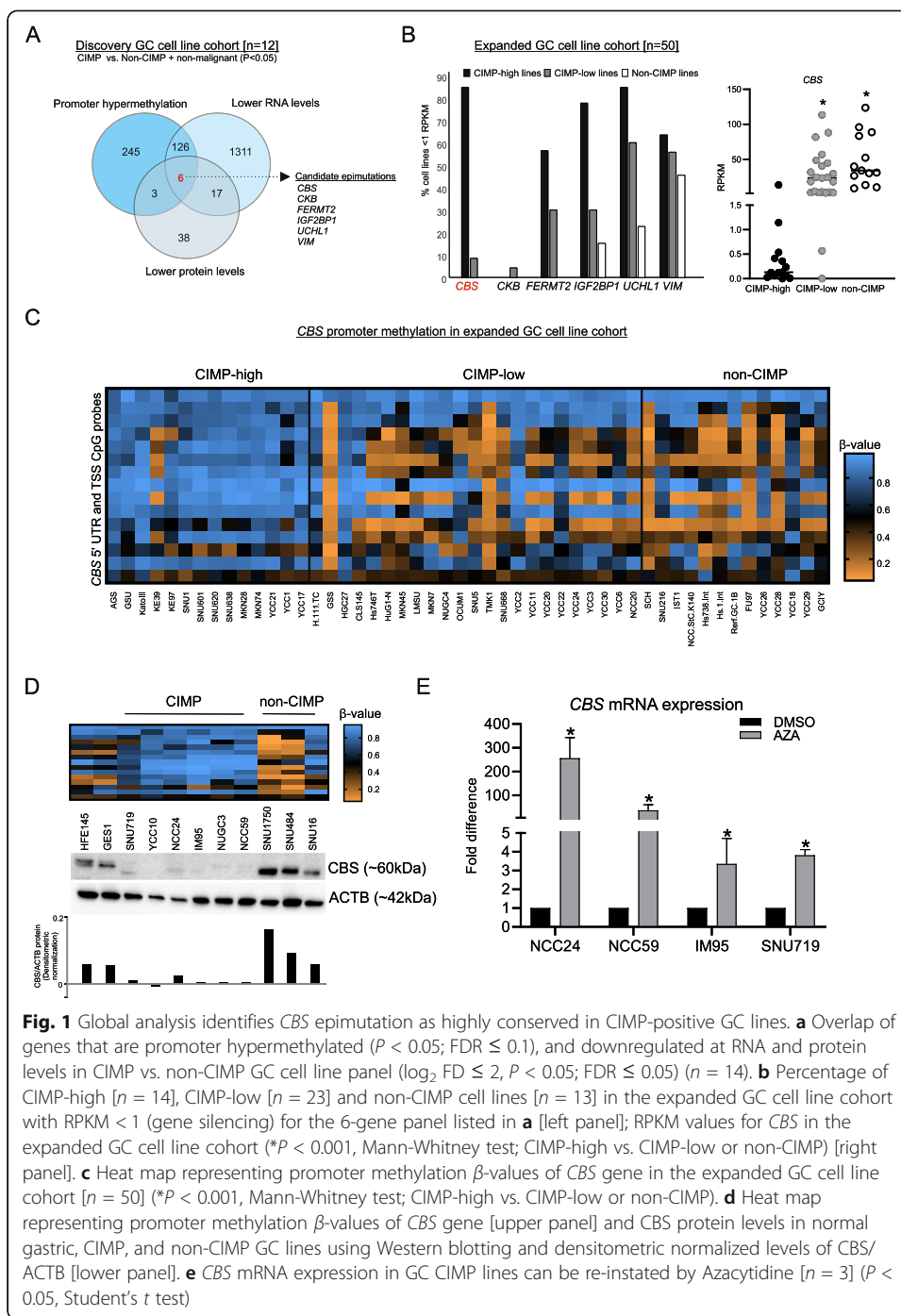
To identify conserved GC CIMP epimutations, we assembled a panel of 14 gastric cell lines based on previous literature (“Methods”) that comprised 3 gastric/MSI CIMP (NUGC3, NCC59, and IM95), 3 EBV-CIMP (SNU719, YCC10 and NCC24), 6 non-CIMP GC cell lines (SNU16, SNU484, SNU1967, SNU1750, NCC19 and MKN1), and

two non-malignant gastric epithelial cell lines (GES1 and HFE145). We verified the CIMP and non-CIMP status of each line based on hierarchical clustering of top variable 10,000 CGI probes derived from Illumina HumanMethylation450 Infinium DNA methylation array data (“Methods,” Additional file 1: Figure S1).

We performed global methylation sequencing, RNA sequencing, and proteomic analysis on the 14-cell line panel. Methylomes were interrogated using MeDIP, a technique that can distinguish 5-methyl cytosine from 5-hydroxy methyl cytosine, a demethylation derivative with different transcriptional associations. Combining the gastric CIMP and EBV-CIMP lines, we identified genes exhibiting both aberrant gain of promoter methylation ( $P < 0.05$ ;  $FDR \leq 0.1$ ) (Additional file 2) and significant downregulation at the RNA (average  $\log_2$  FD  $\leq 2$ ,  $P < 0.05$ ;  $FDR \leq 0.05$ ) (Additional file 3) and protein level (average  $\log_2$  FD  $\leq 2$ ,  $P < 0.05$ ;  $FDR \leq 0.05$ ) (Additional files 4, 5), compared to non-CIMP and non-malignant gastric lines (Fig. 1a). We found a set of 6 candidate epimutations, consisting of *CBS* [cystathionine beta-synthase], *CKB* [creatinine kinase B], *FERMT2* [fermitin family member 2], *IGF2BP1* [insulin like growth factor 2 mRNA binding protein 1], *UCHL1* [ubiquitin C-terminal hydrolase L1], and *VIM* [vimentin] (Fig. 1a).

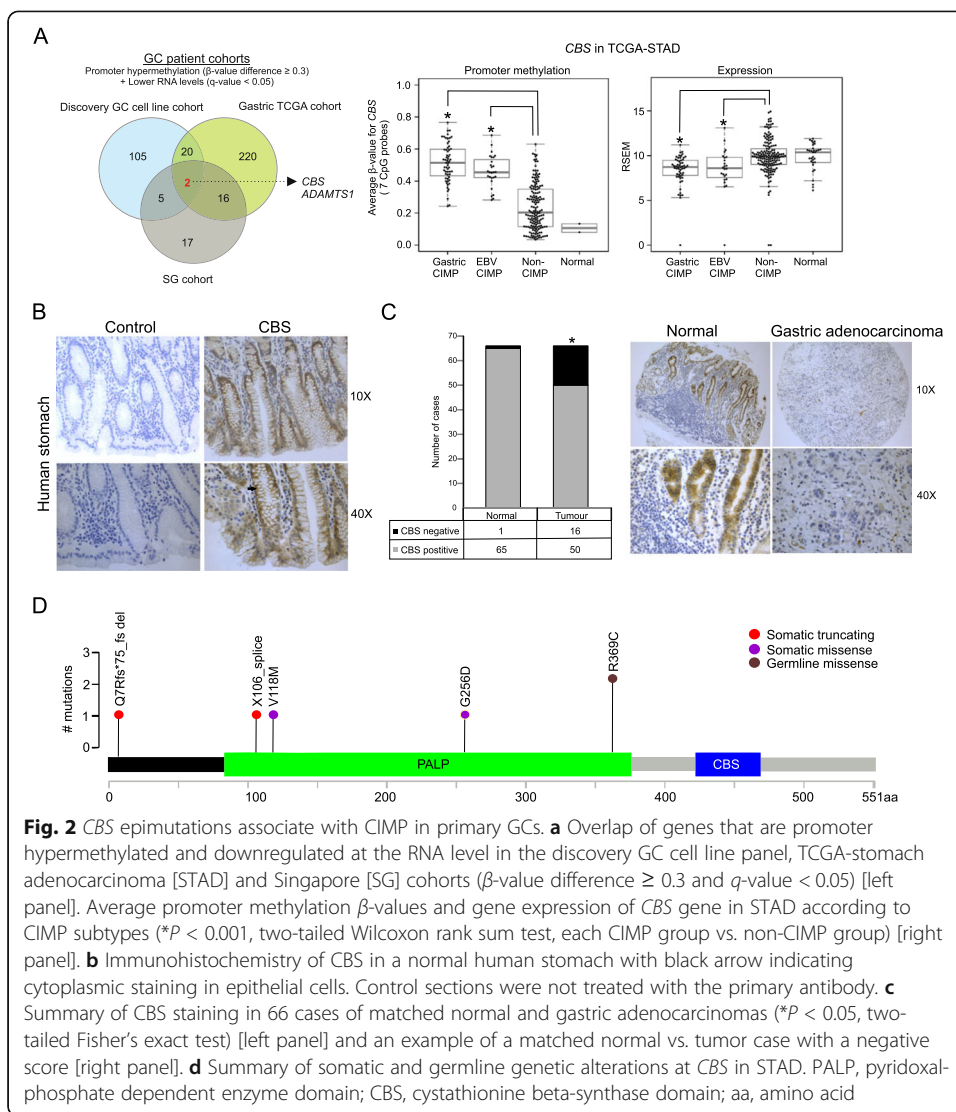
To establish the most conserved CIMP epimutation among these 6 candidates, we assessed RNA levels of these 6 genes in an expanded cohort of 50 GC cell lines subtyped as “CIMP-high” ( $n = 14$ ), “CIMP-low” ( $n = 23$ ), and “non-CIMP” ( $n = 13$ ) (“Methods,” Additional file 1: Figure S1). A recurrence analysis in these 50 GC cell lines highlighted *CBS* transcriptional loss as the most frequent and specific marker in CIMP-high cell lines (12/14, 85.71%) compared to CIMP-low (2/23, 8.69%) and non-CIMP cell lines (0/13, 0%) (Additional file 6), with significant negative correlations to increased *CBS* promoter methylation (Spearman  $r = -0.72$ ,  $P < 0.001$ ) (Fig. 1b,c, Additional file 7). Using quantitative PCR and Western blotting, we orthogonally confirmed that *CBS* mRNA and protein are expressed in normal gastric and non-CIMP cell lines but downregulated in CIMP cell lines (Fig. 1d, Additional file 1: Figure S2A, S3; Refer Fig. S2B for *CBS* antibody validation). Treatment of GC CIMP lines with a DNA demethylating agent (Azacytidine) restored *CBS* mRNA expression (FD = 3.3–255.8 vs. DMSO,  $P < 0.05$ ), reaffirming the negative relationship between *CBS* epimutation and expression (Fig. 1e) [14]. Since hypermethylated promoter loci in GC are frequently targeted by PRC2 which predominantly catalyzes histone H3 trimethylation on lysine 27 [4, 15], we assessed H3K27me3 status at the *CBS* gene promoter in 2 gastric CIMP cell lines. However, we found no enrichment of this histone mark (Additional file 1: Figure S4), reinforcing a stronger role for DNA methylation in *CBS* silencing, likely in concert with other repressive histone marks (e.g., H3K9me3).

Although GC cell lines offer the key advantage of representing a pure population of tumor gastric epithelial cells, we next attempted to validate our findings in primary GCs to eliminate confounders that may have arisen due to in vitro effects. We used a similar approach to analyze methylation array ( $\beta$ -value difference  $\geq 0.3$ ) and gene expression data ( $q$ -value  $< 0.05$ ) in two GC patient cohorts (gastric TCGA [10] and Singapore Cohort [4],  $n = 467$  patients). Despite platform differences (“Methods”), this integrative analysis yielded a common signature of 18 genes that included *CBS* (Fig. 2a, Additional file 8). Of the remaining 5 candidate epimutations, *CKB* and *FERMT2* were also significantly associated with CIMP in the gastric TCGA cohort. We verified expression of *CBS* in normal human stomach using quantitative PCR and immunohistochemistry (Fig. 2b, Additional file 1: Figure S2C-D), revealing predominant cytoplasmic staining of *CBS* in the gastric epithelium, consistent with a previous report in mice



**Fig. 1** Global analysis identifies *CBS* epimutation as highly conserved in CIMP-positive GC lines. **a** Overlap of genes that are promoter hypermethylated ( $P < 0.05$ ;  $FDR \leq 0.1$ ), and downregulated at RNA and protein levels in CIMP vs. non-CIMP GC cell line panel ( $\log_2$  FD  $\leq 2$ ,  $P < 0.05$ ;  $FDR \leq 0.05$ ) ( $n = 14$ ). **b** Percentage of CIMP-high [ $n = 14$ ], CIMP-low [ $n = 23$ ] and non-CIMP cell lines [ $n = 13$ ] in the expanded GC cell line cohort with RPKM < 1 (gene silencing) for the 6-gene panel listed in **a** [left panel]; RPKM values for *CBS* in the expanded GC cell line cohort ( $*P < 0.001$ , Mann-Whitney test; CIMP-high vs. CIMP-low or non-CIMP) [right panel]. **c** Heat map representing promoter methylation  $\beta$ -values of *CBS* gene in the expanded GC cell line cohort [ $n = 50$ ] ( $*P < 0.001$ , Mann-Whitney test; CIMP-high vs. CIMP-low or non-CIMP). **d** Heat map representing promoter methylation  $\beta$ -values of *CBS* gene [upper panel] and CBS protein levels in normal gastric, CIMP, and non-CIMP GC lines using Western blotting and densitometric normalized levels of CBS/ACTB [lower panel]. **e** *CBS* mRNA expression in GC CIMP lines can be re-instated by Azacytidine [ $n = 3$ ] ( $P < 0.05$ , Student's *t* test)

[16]. We further tested for CBS protein expression in a tissue microarray comprising 66 cases of matched normal and GCs. A significant proportion of cancer cases exhibited reduced CBS staining in the malignant gastric epithelium (24.2%) compared to matched normal (1.5%) ( $P < 0.05$ ) (Fig. 2c). Unfortunately, associations with CIMP status could not be performed due to lack of methylation data for this GC cohort. These results support *CBS* epimutation as the most recurrent epimutation associated with GC CIMP in both cell lines and primary tumors [14].



**Fig. 2** CBS epimutations associate with CIMP in primary GCs. **a** Overlap of genes that are promoter hypermethylated and downregulated at the RNA level in the discovery GC cell line panel, TCGA-stomach adenocarcinoma [STAD] and Singapore [SG] cohorts ( $\beta$ -value difference  $\geq 0.3$  and  $q$ -value  $< 0.05$ ) [left panel]. Average promoter methylation  $\beta$ -values and gene expression of CBS gene in STAD according to CIMP subtypes ( $*P < 0.001$ , two-tailed Wilcoxon rank sum test, each CIMP group vs. non-CIMP group) [right panel]. **b** Immunohistochemistry of CBS in a normal human stomach with black arrow indicating cytoplasmic staining in epithelial cells. Control sections were not treated with the primary antibody. **c** Summary of CBS staining in 66 cases of matched normal and gastric adenocarcinomas ( $*P < 0.05$ , two-tailed Fisher’s exact test) [left panel] and an example of a matched normal vs. tumor case with a negative score [right panel]. **d** Summary of somatic and germline genetic alterations at CBS in STAD. PALP, pyridoxal-phosphate dependent enzyme domain; CBS, cystathionine beta-synthase domain; aa, amino acid

We also interrogated the gastric TCGA cohort for somatic and germline mutations in the CBS gene. Analysis of 243 GCs identified 6 cases—4 cases with CBS somatic mutations all of which were predicted to be truncating or pathogenic (FATHMM score  $> 0.9$ ), and 2 cases with predicted deleterious germline CBS mutations (Fig. 2d, Additional file 9). Mutated cases were heterozygous, with a low prevalence (6/243, 2.46% frequency), co-occurred with gastric CIMP and MSI (Fisher’s exact test,  $P = 0.001$ ), and exhibited promoter hypermethylation ( $\beta$ -value range 0.59–0.79), suggesting that genetic mutations are much less frequent than epimutations at the CBS locus and may or may not represent passenger events owing to MSI.

### CBS epimutations associate with CIMP in multiple cancer types and premalignant intestinal metaplasia

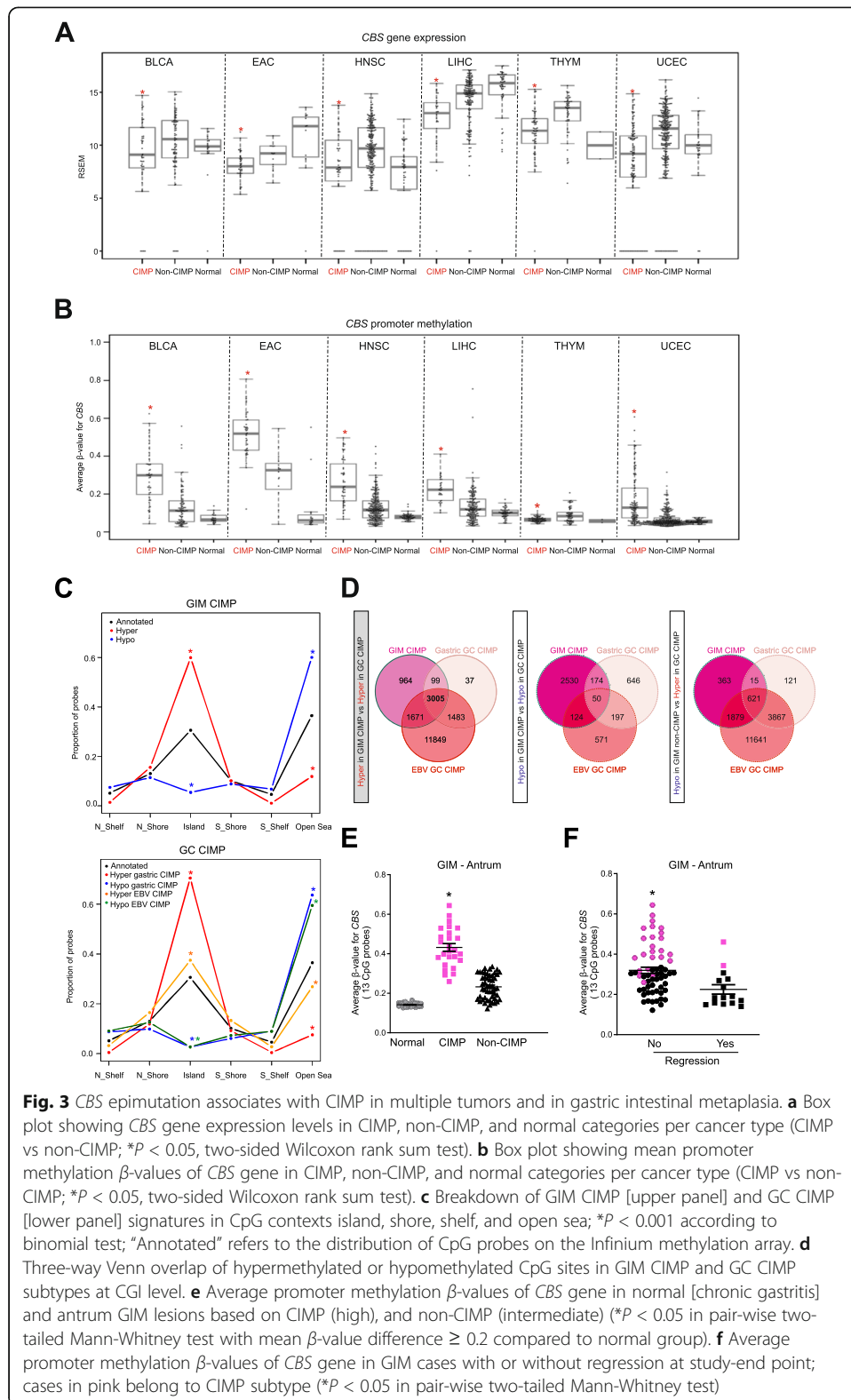
We then investigated if CBS epimutations are associated with CIMP in other cancers. Using TCGA-generated datasets and tumor categorizations of CIMP that has been

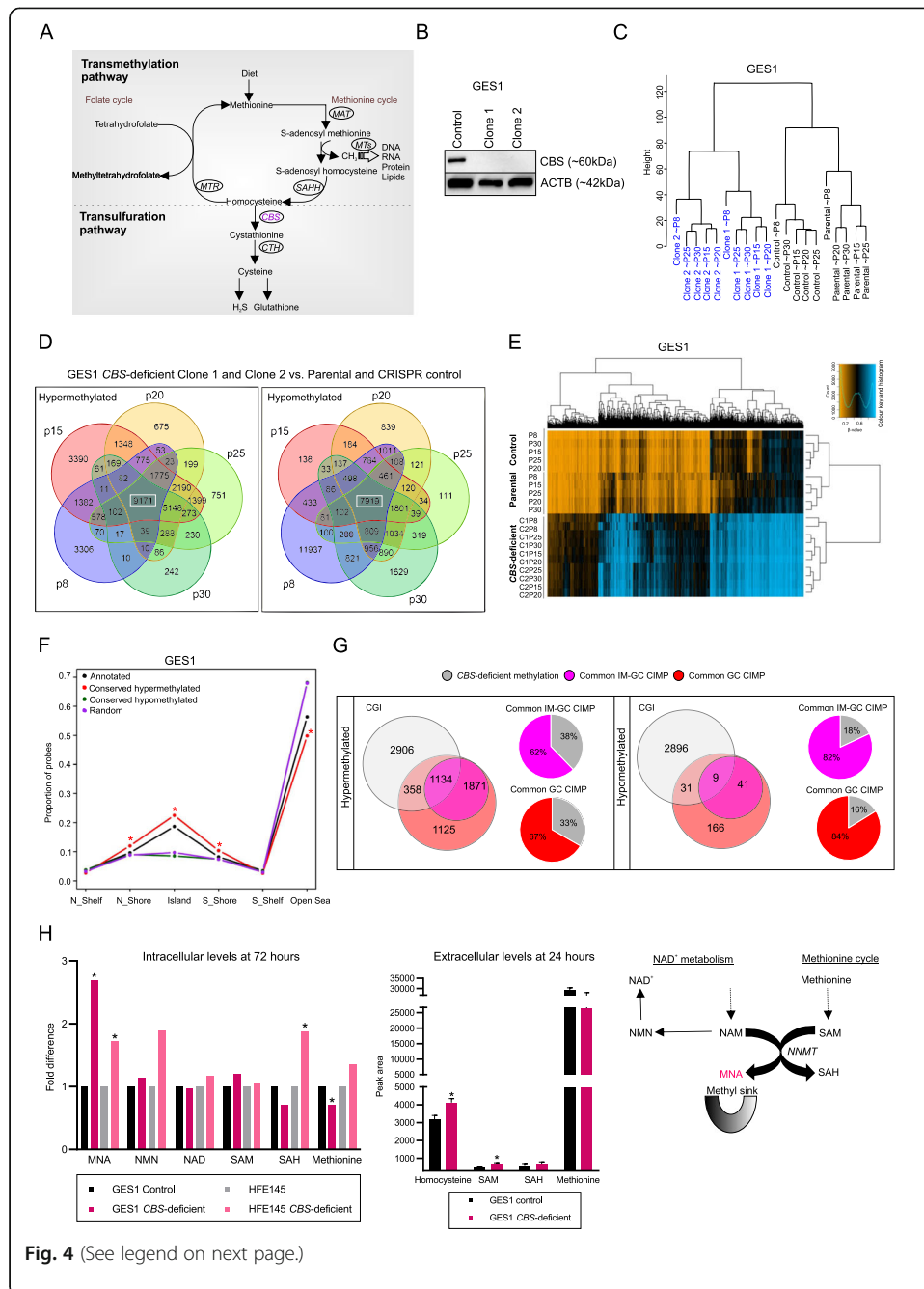
previously described (“Methods”), we analyzed 22 different cancer types performing *CBS* differential gene expression analysis between most “hypermethylated/CIMP” vs. other methylation categories (Wilcoxon rank sum test,  $P < 0.05$ , “Methods”). We identified 6 cancers displaying significant downregulation of *CBS* in CIMP vs non-CIMP tumors, including bladder urothelial carcinoma [BLCA] ( $P = 0.02$ ), esophageal adenocarcinoma [EAC] ( $P = 0.01$ ), head and neck squamous cell carcinoma [HNSC] ( $P = 0.006$ ), liver hepatocellular carcinoma [LIHC] ( $P = 3.5 \times 10^{-5}$ ), thymoma [THYM] ( $P = 3.2 \times 10^{-6}$ ), and uterine corpus endometrial carcinoma [UCEC] ( $P = 3.6 \times 10^{-11}$ ) (Fig. 3a). Differential methylation analysis in these 6 cancers further confirmed significant *CBS* promoter hypermethylation in 5/6 of these cancers including BLCA ( $P = 1.1 \times 10^{-10}$ ), EAC ( $P = 6.1 \times 10^{-8}$ ), HNSC ( $P = 4.1 \times 10^{-13}$ ), LIHC ( $P = 2.1 \times 10^{-9}$ ), and UCEC ( $P = 1.7 \times 10^{-22}$ ) (Fig. 3b, Additional file 1: Figure S5), with significant inverse correlations to gene expression (Spearman: BLCA  $r = -0.21$ ,  $P = 0.01$ ; EAC  $r = -0.46$ ,  $P = 2.1 \times 10^{-4}$ ; HNSC  $r = -0.28$ ,  $P = 4.1 \times 10^{-7}$ , LIHC  $r = -0.33$ ,  $P = 9.8 \times 10^{-8}$ ; UCEC  $r = -0.25$ ,  $P = 1.9 \times 10^{-7}$ , Additional file 1: Figure S6) indicating that *CBS* epimutations are associated with CIMP in other tumor types besides GC.

Gastric intestinal metaplasia [GIM] is a premalignant condition of the stomach associated with an estimated annual GC risk of  $\sim 0.25\%$  [17]. We recently described the existence of CIMP in GIM [15]. Like GC, GIM CIMP is also significantly enriched for hypermethylated CGIs ( $P < 0.001$ ) (Fig. 3c). To assess conserved hypermethylation between GIM and GC CIMP, we compared CGIs (with a minimum of 1 CpG probe overlap) between GIM CIMP, GC CIMP, and non-CIMP groups ( $\beta$ -value difference  $\geq 0.1$ , FDR  $\leq 0.05$ ;  $P < 1 \times 10^{-5}$ , bootstrapping using 100,000 iterations) and identified a “GIM-GC common CIMP signature” of 3005 CGIs (Fig. 3d; Additional file 10). Notably, the GIM-GC common CIMP signature again included aberrant *CBS* promoter hypermethylation (average  $\beta$ -value difference = 0.28 between GIM CIMP compared to chronic gastritis in antrum,  $P < 0.05$ ), indicating its early occurrence in the pre-cancerous GC cascade (Fig. 3e). Over a 5-year longitudinal period, GIMs with increased *CBS* promoter methylation were significantly associated with clinical persistence and lack of regression ( $P < 0.05$ ) (Fig. 3f). Similar to GIM and GC, EAC is known to display CIMP and considered to arise from intestinal metaplastic lesions in the esophagus (“Barrett’s esophagus” [BE]). We investigated the methylation status of the *CBS* promoter in BE premalignant lesions using a publicly available dataset [18]. Echoing our observations in GIM, the *CBS* promoter was significantly hypermethylated in BE and EAC but not in control tissues ( $P < 0.05$ ) (Additional file 1: Figure S7).

#### Genetic ablation of *CBS* in normal gastric cells results in abnormal DNA methylation

*CBS* encodes an enzyme occupying a unique biochemical position in the methionine and homocysteine metabolic pathways (Fig. 4a). Homocysteine is converted into methionine and subsequently *S*-adenosyl methionine [SAM], the universal methyl donor for DNA, RNA, histones, and lipids. Because *CBS* can break down homocysteine into cystathionine, we reasoned that *CBS* loss may lead to homocysteine accumulation and disrupt normal methyl group homeostasis, and potentially implicate *CBS* loss-of-function epimutations as more than a marker of GC CIMP. This is also supported by the fact that *CBS*-deficient mice display significant plasma homocysteinemia (where







(See figure on previous page.)

**Fig. 4** Knocking out CBS alters the DNA methylation landscape in normal gastric epithelial cells. **a** Biochemical position of CBS in the transsulfuration and transmethylation cycles. **b** CBS protein levels in GES1 CRISPR control and CBS-deficient CRISPR clones. **c** Cluster dendrogram of methylation data indicating branching of GES1 CBS-deficient clones from GES1 parental and CRISPR control cells. **d** Venn diagram of differential methylation analysis of GES1 CBS-deficient clone 1 and clone 2 compared to parental and CRISPR control cells. **e** Unsupervised clustering of conserved hypermethylated 9171 CpG sites in GES1 CBS-deficient clones [clone 1, C1; clone 2, C2] across passages. **f** Breakdown of the GES1 CBS-deficient conserved signature in CpG contexts island, shore, shelf, and open sea; \* $P < 0.001$  according to binomial test; “Annotated” refers to the distribution of CpG probes on the Infinium methylation EPIC array, “Random” refers to the sum of all differentially methylated CpG probes in random pairing of samples. **g** Comparison of CBS-deficient conserved signatures in GES1 or HFE145 with the “GIM-GC common CIMP” and “GC common CIMP” signatures at CpG islands indicate a larger overlap for hypermethylated (33–38%) events compared to hypomethylation events (16–18%);  $P < 1 \times 10^{-4}$ , bootstrapping using 10,000 samples. **h** Graph showing fold difference of intracellular levels of metabolites in CBS-deficient cells ( $n = 4-5$ ) compared to controls ( $n = 5$ ) at 72 h, 1-methyl nicotinamide [MNA] (and related metabolites, nicotinamide mononucleotide [NMN], nicotinamide adenine dinucleotide [NAD], S-adenosyl methionine [SAM], S-adenosyl homocysteine [SAH], and methionine, \* $P < 0.05$  (left panel). Graph showing peak area of extracellular levels of homocysteine, SAM, SAH, and methionine in CBS-deficient cells ( $n = 5$ ) compared to controls ( $n = 7$ ) at 24 h, \* $P < 0.05$  (middle panel). Biochemical reaction showing conversion of nicotinamide [NAM] to MNA using a methyl group from SAM, catalyzed by nicotinamide N-methyltransferase [NNMT] (right panel). CBS, cystathionine beta-synthase; CTH, cystathionine gamma-lyase; MTR, 5-methyltetrahydrofolate-homocysteine methyltransferase; MAT, Methionine adenosyltransferase; MTs, Methyl transferases

excess cellular homocysteine is exported), and tissue-specific metabolic defects in the methionine pathway [19]. To investigate this possibility, we employed CRISPR genome editing to generate CBS loss-of-function models in a non-malignant gastric epithelial cell line (GES1), using two different guide RNAs (Fig. 4b; Additional file 1: Figure S3).

To study DNA methylation in GES1 CBS-deficient clones, we used Infinium methylation EPIC arrays to assay GES1 parental, CRISPR control with a non-targeting sequence, CBS-deficient clone 1, and CBS-deficient clone 2 across 5 different passage numbers (~P8 to P30, culture time  $\leq 3$  months). Hierarchical clustering of GES1 parental, CRISPR control, and CBS-deficient clones indicated a clear divergence of the latter, suggesting that CBS loss induces genome-wide methylome changes (Fig. 4c). To minimize clonal effects, we performed differential methylation analysis of GES1 CBS-deficient clones ( $n = 10$ ) compared to parental and CRISPR control cells ( $n = 10$ ), such that a minimum  $\beta$ -value difference  $\geq 0.2$  was present in both clones compared to all controls at each passage number. This analysis indicated a conserved signature of 17,090 CpG sites (9171 hypermethylated and 7919 hypomethylated CpG sites; Fig. 4d,e, Additional file 11), which was absent in random pairings of samples (Additional file 1: Figure S8). Importantly, this conserved hypermethylation exhibited features of in vivo CIMP, including enrichment for CGIs and depletion of open sea regions ( $P < 0.001$ ), and enrichment of PRC2 targets ( $P < 0.001$ ) (Fig. 4f, Additional file 1: Figure S9F). Analysis of hypermethylated CpG sites with gene expression indicated passenger and non-passenger methylation events, with a distinct subset of genes downregulated ( $FD \leq 0.6$ ) in both GES1 CBS-deficient clones in the latter category (6.8%, 76/1155 genes) (Additional file 12). These included cancer epigenetic targets previously implicated in tumor suppressive activity such as ANXA6 [annexin A6] [20], VANGL2 [VANGL planar cell polarity protein 2] [21], BIN1 [bridging integrator 1] [22], and CREB3L1 [cAMP responsive element binding protein 3 like 1] [23].

To rule out cell-line-specific effects, we generated CBS-deficient clones in another non-malignant gastric epithelial cell line (HFE145), performing a similar methylation analysis

at two different passage numbers (P5 and P23) (Additional file 1: Figures S9A, S3). Differential methylation analysis of HFE145 *CBS*-deficient clones compared to parental and CRISPR control cells yielded a conserved signature of 6886 CpG sites across both passages (Additional file 1: Figure S9B-C, Additional file 13). Although the magnitude of change in HFE145 *CBS*-deficient clones was milder than that of GES1 *CBS*-deficient clones, the conserved hypermethylation was still enriched for CGIs and depleted for open sea regions ( $P < 0.001$ ), and enriched for PRC2 targets ( $P < 0.001$ ) (Additional file 1: Figure S9D-F). Analysis of hypermethylated CpG sites with gene expression again indicated passenger and non-passenger methylation events, with 6.1% (29/468 genes) of genes downregulated ( $FD \leq 0.6$ ) in both HFE145 *CBS*-deficient clones (Additional file 14).

To evaluate the concordance of DNA hypermethylation targets derived in vitro due to *CBS* deficiency and in vivo CIMP profiles, we compared the signatures observed in *CBS*-deficient clones (GES1 and/or HFE145) with the previously identified “GIM-GC common CIMP signature.” We observed a significant  $\sim 38\%$  overlap at CGIs ( $P < 1 \times 10^{-4}$ , bootstrapping using 10,000 samples, Fig. 4g, Additional file 15) that included candidate tumor suppressor *ROBO1* [roundabout guidance receptor 1] [24, 25] that is known to be promoter hypermethylated and inactivated in multiple cancers.

Since we observed a DNA methylation phenotype in *CBS*-deficient cells, we also asked if *CBS* deficiency caused metabolic changes in gastric epithelial cells, particularly in metabolites related to the methylation pathway. We profiled  $\sim 244$ – $327$  compounds (polar and lipid) in cellular extracts from GES1 and HFE145 cell line models at 72 h and performed differential analyses of *CBS*-deficient clones compared to parental and CRISPR control cells (Additional file 16, “Methods”). We identified 8 metabolites altered in *CBS*-deficient clones of both cell lines (4 increased, 4 decreased). Interestingly, these included increased levels of “1-methyl nicotinamide” [MNA] ( $FD$  1.7–2.6;  $P < 0.05$ ), a metabolite linked to the SAM cycle and considered a “methyl sink” (Fig. 4h, Additional file 1: Figure S10A), where excess methyl groups from SAM are transferred to nicotinamide (Fig. 4h, left and right panels) [26–28]. The mRNA expression of nicotinamide N-methyltransferase enzyme that catalyzes this reaction remained unchanged in *CBS*-deficient clones (Additional file 1: Figure S10B), implying that increased MNA levels reflect disturbances in methyl group homeostasis rather than enzyme expression levels. In addition, we also observed multiple triglycerides and phospholipids exhibiting significant changes, consistent with a prior study in *CBS*-deficient mice that reported that *CBS* loss alters lipid metabolism [29] (Additional file 1: Figure S10C). Of note, intracellular homocysteine could not be detected in this global analysis, likely due to the sensitivity of platform to detect low levels at 72 h [30]. Since excess homocysteine is usually exported out of the cell, we then measured homocysteine, SAM, S-adenosyl homocysteine [SAH], and methionine in culture media from the GES1 cell line model at 24 h using targeted assays (“Methods”). We found significant increases in homocysteine ( $FD$  1.28;  $P < 0.05$ ) and SAM ( $FD$  1.50;  $P < 0.05$ ), but not SAH and methionine in *CBS*-deficient cells (Fig. 4h, middle panel), further suggestive of an imbalance in the methyl group cycle. Taken collectively, we find that loss of *CBS* in two independent normal gastric epithelial cell lines was associated with an abnormal methylation landscape that overlapped with in vivo CIMP patterns and an altered metabolic profile.

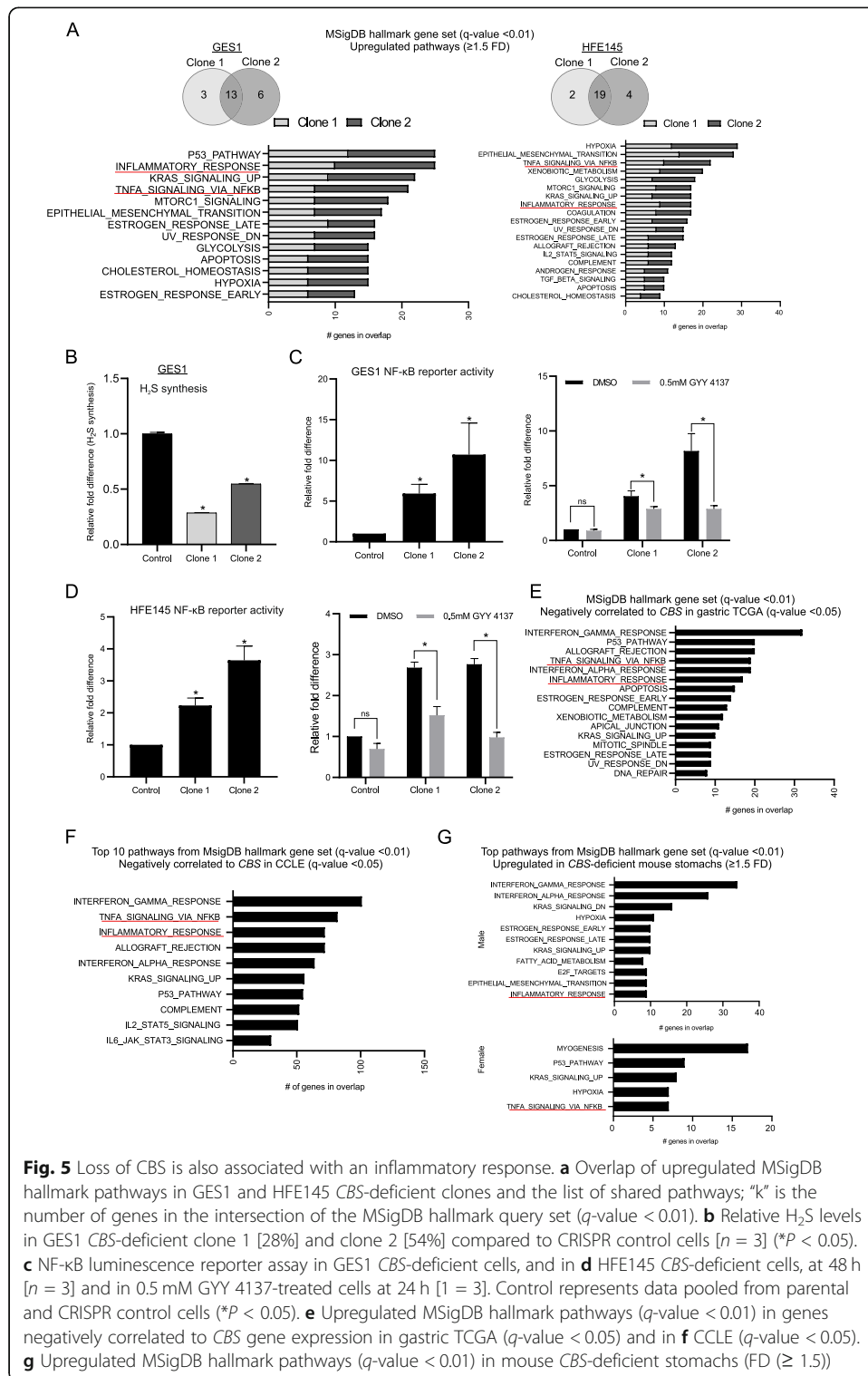
### **CBS loss is also associated with an inflammatory state**

Unexpectedly, besides DNA methylation alterations, RNA sequencing analysis of both GES1 and HFE145 *CBS*-deficient clones compared to parental and CRISPR control cells revealed an upregulation of inflammation-related pathways, such as “*Inflammatory response*” and “*TNFA signalling via NF- $\kappa$ B*” ( $q$ -value < 0.01) (Fig. 5a, Additional files 17 and 18). To reaffirm this finding, we measured NF- $\kappa$ B activity (pro-inflammatory complex) in GES1 and HFE145 *CBS*-deficient clones using a reporter assay and observed significantly higher activity in these cells (5.9–10.7 FD in GES1; 2.2–3.6 FD in HFE145, compared to parental and control CRISPR cells,  $P$  < 0.05) (Fig. 5c, left panel; Fig. 5d, left panel).

To understand this phenomenon in greater detail, we revisited the *CBS* metabolic network and noted that besides being linked to transmethylation, *CBS* also catalyzes the first step of the transsulfuration pathway, leading to the generation of hydrogen sulfide [ $\text{H}_2\text{S}$ ] and glutathione [31]. Consistent with the functional loss of *CBS*, both GES1 *CBS*-deficient clones exhibited significantly lower  $\text{H}_2\text{S}$  levels (0.28–0.54 FD compared to CRISPR control cells,  $P$  < 0.05) (Fig. 5b).  $\text{H}_2\text{S}$  is a gasotransmitter, and the *CBS*- $\text{H}_2\text{S}$  axis has been implicated in multiple physiological processes including inflammation, vasodilation, and cell metabolism, in part through  $\text{H}_2\text{S}$ 's function of “sulfhydration”—a post-translational modification on proteins [32–34]. We thus reasoned that the induction of inflammatory pathways in *CBS*-deficient cells could in part be caused by reductions in  $\text{H}_2\text{S}$  levels [35]. Treatment with a  $\text{H}_2\text{S}$  slow-release donor GYY 4137 at 0.5 mM for 24 h significantly ameliorated NF- $\kappa$ B activity in *CBS*-deficient cells (1.4–2.8 FD in GES1; 0.3–0.5 FD in HFE145, compared to DMSO control,  $P$  < 0.05) (Fig. 5c, right panel; Fig. 5d, right panel), supporting the idea that loss of  $\text{H}_2\text{S}$  in *CBS*-deficient cells contributes to a portion of the inflammatory response.

To validate these in vitro findings in primary GCs and other data sets, we identified genes that were negatively correlated in expression to *CBS* in the gastric TCGA cohort ( $q$ -value < 0.05). We found that these genes significantly overlapped with inflammation-related pathways in the MSigDB hallmark gene sets that included “*Inflammatory response*” and “*TNFA signalling via NF- $\kappa$ B*” ( $q$ -value < 0.01) (Fig. 5e, Additional file 19). Likewise, in BLCA, ESCA, HNSC, and LIHC, genes that were negatively correlated in expression to *CBS* ( $q$ -value < 0.05) significantly overlapped with “*Inflammatory response*” pathways ( $q$ -value < 0.01) (Additional file 20). A similar gene expression analysis from CCLE [ $n$  = 1739 cell lines [36]] again confirmed that genes negatively correlated in expression to *CBS* ( $q$ -value < 0.05) overlapped with inflammation-related pathways including “*Inflammatory response*” and “*TNFA signalling via NF- $\kappa$ B*” ( $q$ -value < 0.01) (Fig. 5f, Additional file 21).

Finally, we explored the consequences of *CBS* deficiency in vivo using *Tg-hCBS Cbs*<sup>-/-</sup> knockout mice (“Methods”) [37]. Hematoxylin and eosin staining of *Cbs*<sup>-/-</sup> stomachs ( $n$  = 11; aged 2.1 to 11.1 months) vs. *Cbs*<sup>+/+</sup> stomachs ( $n$  = 6; aged 2.8–8.1 months) showed no overt histological defects or neoplasia at this age (data not shown). We performed RNA sequencing in age-matched *Cbs*<sup>-/-</sup> ( $n$  = 2; 1 male, 1 female) and *Cbs*<sup>+/+</sup> ( $n$  = 2; 1 male, 1 female) stomachs for differential gene expression analysis (FD  $\geq$  1.5 and FD  $\leq$  0.6) (Additional file 22). As expected, *CBS* mRNA levels were lower in *Cbs*<sup>-/-</sup> stomachs (FD = 0.08–0.14 compared to *Cbs*<sup>+/+</sup> stomachs, Additional file 22). Despite low sample numbers, pathway analysis using PANTHER (FDR < 0.05) identified



**Fig. 5** Loss of CBS is also associated with an inflammatory response. **a** Overlap of upregulated MSigDB hallmark pathways in GES1 and HFE145 CBS-deficient clones and the list of shared pathways; “k” is the number of genes in the intersection of the MSigDB hallmark query set ( $q$ -value  $< 0.01$ ). **b** Relative H<sub>2</sub>S levels in GES1 CBS-deficient clone 1 [28%] and clone 2 [54%] compared to CRISPR control cells [ $n = 3$ ] ( $*P < 0.05$ ). **c** NF- $\kappa$ B luminescence reporter assay in GES1 CBS-deficient cells, and in **d** HFE145 CBS-deficient cells, at 48 h [ $n = 3$ ] and in 0.5 mM GYY 4137-treated cells at 24 h [ $n = 3$ ]. Control represents data pooled from parental and CRISPR control cells ( $*P < 0.05$ ). **e** Upregulated MSigDB hallmark pathways ( $q$ -value  $< 0.01$ ) in genes negatively correlated to CBS gene expression in gastric TCGA ( $q$ -value  $< 0.05$ ) and in **f** CCLE ( $q$ -value  $< 0.05$ ). **g** Upregulated MSigDB hallmark pathways ( $q$ -value  $< 0.01$ ) in mouse CBS-deficient stomachs (FD  $\geq 1.5$ )

“circadian regulation of gene expression” as the top upregulated pathway, consistent with a recent study reporting a distinct role for CBS in circadian function [38]. More pertinent to this study, several immune-related processes were also significantly upregulated including “complement activation” and “defense response to bacterium,”

indicative of an inflammatory state in *CBS*-deficient stomachs (Additional file 23) and concurring with our earlier observations. In addition, pathway analysis using MSigDB hallmark gene sets also indicated “*Inflammatory response*” and “*TNFA signalling via NF- $\kappa$ B*” as significantly upregulated ( $q$ -value < 0.01) (Fig. 5g; Additional file 24). A similar PANTHER (FDR < 0.05) pathway analysis of downregulated genes in *Cbs*<sup>-/-</sup> stomachs did not yield significant hits.

## Discussion

Using a multi-platform approach comprising methylation, RNA, and protein datasets to identify “epimutations” in GC, we determined *CBS* as a recurrent and conserved epigenetic target in CIMP tumors, associated with both MSI and EBV GC. In vitro loss of CBS in normal human gastric epithelial cells is associated with both aberrant DNA methylation and an inflammatory response, indicating a novel bi-faceted function to this epimutation. Whether these in vitro phenotypes are observed in mouse models, and/or exacerbate with prolonged periods of time, following a “methylation accumulation model” needs to be ascertained, given that primary lesions also take years to develop. Nevertheless, the dual phenotypes are pertinent given a long-standing link between CIMP lesions and inflammation, with chronic inflammation considered a universal accelerator of DNA hypermethylation [39–41]. Notably, pan-cancer analysis indicated that these associations were observed in other tumor types beyond GC, implying a broader role for *CBS* epimutation in shaping tumor biology. The tissue-independent mechanism leading to epigenetic targeting of *CBS* is still unknown. It may be triggered either by a fundamental cellular insult, dysregulation of factors impacting *CBS* expression including non-coding RNAs [42], or perhaps arises as an adaptation to disturbances in biochemically connected pathways such as methionine levels [43, 44]. It is relevant to point out that prior cancer studies have reported both loss and overexpression of the *CBS* gene, relating to tumor suppressive and oncogenic properties respectively, with the latter associated with an abnormal bioenergetic phenotype [45]. We also observe high levels of CBS expression in a proportion of GC cell lines that tend to be non-CIMP (Fig. 1b), suggesting that *CBS* gene dosage is a key factor to maintain a normal metabolic state, and its expression distinctly varies with tumor subtype.

We demonstrated that *CBS* deficiency in normal gastric cell lines resulted in increased homocysteine, SAM, and MNA (“methyl sink”), findings also seen in other models where methylation patterns are abnormal and indicative of altered methyl group homeostasis [26, 27]. While our study has not identified the exact mechanism by which *CBS* deficiency results in abnormal DNA methylation, we speculate that the phenomenon may be complex, involving metabolic perturbations in the transmethylation cycle (e.g., other classes of methyl group acceptors such as histones, RNA, lipids), and perhaps also the transsulfuration pathway (e.g., H<sub>2</sub>S, glutathione, inflammatory changes) since several of these components have been previously identified as modifiers of DNA methylation, through direct and indirect mechanisms [46–51]. Further functional studies that mimic the metabolic states associated with *CBS* deficiency (e.g., homocysteine exposure) and culture manipulations are required to identify the precise metabolic sequelae leading to altered methylation. For example, it will be useful to assess metabolic changes in *CBS*-deficient cells grown in media with lower methionine levels, since normal media is considered to contain supraphysiological levels of

methionine and could thus blunt the severity of phenotypes assessed [26]. Interestingly, recent studies in colorectal cancer implicated the process of ageing as a key contributor to CIMP, where spontaneous ageing of normal colon organoids led to a CIMP-like state with enhanced susceptibility to transformation by *BRAF*<sup>V600E</sup> mutations, in turn enhancing the CIMP phenotype [8]. A pan-cancer analysis also found that a majority of CIMP tumors have increased DNA methylation drift, also associated with ageing [52]. Of note, *CBS* downregulation was observed in gene expression signatures associated with methylomic drift by two independent studies in esophageal and colorectal cancers [53, 54]. *CBS* has also been directly connected to ageing in fruit flies, wherein *CBS* overexpression increased lifespan [55]. Whether age-related processes play a causative or consequential role to *CBS* epimutation also needs to be examined. A limitation of our study is that the current experimental design may not accurately model fine-scale methylation changes across time, as this will involve a larger set of biological replicates/passage with extended culture times.

Our findings in *CBS*-deficient gastric cells that lower H<sub>2</sub>S levels are associated with increased NF-κB activity, draws from multiple studies in other cell line and animal model systems that have investigated this mechanistically [35, 56–58]. These previous studies have shown the anti-inflammatory function of *CBS* loss and H<sub>2</sub>S to be multifaceted, involving the following: blocking NF-κB p65 activation, regulating cellular redox status, DNA damage, stabilization of hypoxia-inducible factor (HIF)-1α, regulating immune cell migration, mitochondrial electron transport, and apoptosis. At the molecular level, H<sub>2</sub>S's role in “sulfhydration,” has been linked to some of these biological effects [59]. Future work in the *CBS*-deficient in vitro model will systematically investigate the role of these molecular mechanisms in inducing inflammatory responses. Preliminary investigations in *CBS*-deficient mouse stomachs indicated upregulation of several immune-related processes. In line with these findings, previous studies have reported an inflammatory phenotype in bone, liver, and kidney of *CBS*-deficient mice [34, 60]. Of significance, inflammatory markers of osteoporosis in *CBS*-deficient mice could be dampened using H<sub>2</sub>S donor treatment, synchronous with our data that showed a significant rescue of NF-κB activation in *CBS*-deficient human gastric cells. A similar protective effect of H<sub>2</sub>S donors have also been reported in lipopolysaccharide-induced inflammation, rodent models of gastric ulcer, and BE [33, 61, 62], and in preventing viral infections [63]. Based on these collective findings, we think that an in-depth characterization of *CBS*-deficient and GIM mouse models to malignancy risk, inflammation, and H<sub>2</sub>S releasing therapeutics [64] is further warranted.

Finally, a common CIMP signature between GC and GIM CIMPs is reminiscent of findings from other premalignant settings (e.g., BE and EAC) [18] and supports the “DNA methylation accumulation model.” It also prompts assessment of the existence of CIMP and *CBS* epimutations in preceding lesions to address the chronology of molecular events. Even so, *CBS* promoter hypermethylation was significantly associated with the clinical persistence of GIM lesions. We hypothesize that loss of *CBS* in early gastric lesions could be one of the mechanisms by which abnormal DNA methylation and inflammatory state is induced, which over several years and compounded by other risk factors, ultimately leads to accrual of pro-tumorigenic DNA methylation changes and unresolved chronic inflammation (Additional file 1: Figure S11). It is also possible that *CBS* epimutation likely precedes EBV onset since GIMs are not clonally infected,

and perhaps may even contribute to lowered cellular defense to viral infection. It will be valuable to test these ideas using clinical samples and appropriate mouse models.

## Conclusion

In summary, our study discovered a novel association between *CBS* epimutations and CIMP subtype in GC, with in vitro models of *CBS* deficiency resulting in abnormal DNA methylation and inflammatory response. These findings provide new insights on the CIMP epigenetic class and re-emphasize epimutations as key players in gastric tumor biology.

## Materials and methods

### Cell lines and primary tissue

Cell lines were purchased from Japan Health Science Research Resource Bank (IM95, NUGC3, and MKN1), ATCC (SNU16, AGS) and Korean Cell Line Bank (NCC59, NCC24, SNU1967, SNU719, SNU484, NCC19, and SNU1750). YCC10 and YCC11 (Yonsei Cancer Centre in Seoul, South Korea), GES1 (Dr. Alfred Cheng, Chinese University of Hong Kong), and HFE145 (Dr. Hassan Ashktorab, Howard University) were kind gifts. Cell lines were authenticated using Short Tandem Repeat profiling using ANSI/ATCC ASN-0002-2011 guidelines and tested Mycoplasma negative according to the MycoAlert Mycoplasma Detection Kit (Lonza). Normal human stomach antrum and fundus tissue slides were purchased from Novus Biologicals (NBP2-30203, NBP2-30204). Human tissue microarray slides containing matched normal and gastric cancer cases, and human control tissue sections (liver and muscle) were provided and processed by SingHealth Advanced Molecular Pathology Laboratory, Singapore.

### GC cell line panel and subtyping

EBV and MSI positive and negative cell lines ( $n = 12$ ) for the GC discovery cell line panel were selected based on prior literature [65–69]. For the DNA methylation subtyping of GC cell lines ( $n = 62$ ), raw intensity data from the Infinium HumanMethylation450 beadchip array were analyzed in R using the minfi package to obtain beta values (ratio of methylation). Samples were normalized using the functional normalization function. CpG sites mapping to sex chromosomes or SNP regions were excluded. Similar to the criteria used in the gastric TCGA study, variations (quantified by standard deviations) of the methylation levels at CpG sites within CpG islands across all samples were calculated, and the top 10,000 sites (Additional file 1: Figure S1) were used for an unsupervised hierarchical clustering using Ward's linkage method and Euclidean distance metrics which identified 4 major branches. We categorized these 4 branches into 3 major methylation groups, which we termed "CIMP-high" (branch 1), "CIMP-low" (branches 3 and 4), and "non-CIMP" (branch 2). Infinium HumanMethylation450 beadchip array data and mutational status of select cancer genes for the GC cell lines was accessed through the Singapore Gastric Cancer Consortium (<http://www.sgcc.sg/research/unique-resources>).

### MeDIP-sequencing and analysis

DNA was sonicated using COVARIS S2 and peak fragment distribution between 100 and 500 bp was verified on an Agilent Bioanalyzer (Agilent Technologies) using the

DNA1000 chip. Fragmented DNA was end-repaired, dA-tailed, and adapter ligated using NEBNext® DNA Library Prep Master Mix Set for Illumina (E6040). Samples were then spiked with control DNAs that were unmethylated, methylated, and hydroxymethylated (Diagenode C02040010) as a quality control measure. For each sample, input DNA that was not exposed to the primary antibody was included. Adapter-ligated DNA was subjected to immunoprecipitation with a primary monoclonal antibody against 5-methyl cytosine (Diagenode C15200081) using a previously published protocol [70]. Real-time PCR using primers against the spiked DNA controls were performed to verify successful and specific enrichment of methylated DNA (data not shown). Immunoprecipitated samples were amplified using Phusion® High-Fidelity DNA Polymerase (M0530) and NEBNext® Multiplex Oligos for Illumina® (E7335) for 10 cycles. Amplified libraries were run on the Agilent Bioanalyzer using the High sensitivity DNA kit prior to Illumina sequencing using a single-end 100 base pair configuration. MeDIP short reads were aligned to the h19 reference genome and duplicates were removed using SAM tools. Average library size of the 14 MeDIP libraries were ~ 80 million reads. Peaks were called using MACS2 using input control. Differential peak analysis between cell lines was performed using DiffBind (FDR ≤ 0.1) of CIMP vs. non-CIMP and non-malignant cell lines which generated a list of 595 unique genes with gain of promoter methylation (+/- 2 kb TSS) and annotated using GenCODE.

#### **RNA sequencing and analysis**

Total RNA quality check was done using the RNA 6000 Lab Chip Kit on the Agilent Bioanalyzer (Agilent Technologies, Palo Alto, CA). Libraries were prepared using Illumina Tru-Seq Stranded Total RNA with Ribo-Zero Gold kit protocol, according to the manufacturer's instructions (Illumina, San Diego, CA, USA). Library fragment size was determined using the DNA 1000 Kit on the Agilent Bioanalyzer (Agilent Technologies). Libraries were quantified by qPCR using the KAPA Library Quantification Kit (KAPA Biosystems). Libraries were pooled in equimolar and cluster generation was performed on the Illumina cBOT system (Illumina). Sequencing (150 bp pair-end) was performed on the Illumina HiSeq 3000 system at the Duke-NUS Genome Biology Facility, according to the manufacturer's protocol. The paired end reads were aligned to the human reference genome (GRCh37) using Hisat2. Gene expression profile (FPKM or RPKM) for each sample was determined.

#### **Proteomics and analysis**

Each cell line was grown and extracted in quadruplicates using RIPA buffer (Sigma) according to the manufacturer's instructions. In total, 200 µg of protein was used for MS sample preparation. Samples were boiled at 95 °C prior to separation on a 12% NuPAGE Bis-Tris precast gel (Thermo Fisher Scientific) for 15 min at 170 V in 1× MOPS buffer. The gel was fixed using the Colloidal Blue Staining Kit (Thermo Fisher Scientific), and each lane was divided into 2 equal fractions. For in-gel digestion, samples were destained in destaining buffer (25 mM ammonium bicarbonate; 50% ethanol), reduced in 10 mM DTT for 1 h at 56 °C followed by alkylation with 55 mM iodoacetamide (Sigma) for 45 min in the dark. Tryptic digest was performed in 50 mM ammonium bicarbonate buffer with 2 µg trypsin (Promega) at 37 °C overnight. Peptides were



desalted on StageTips and analyzed by nanoflow liquid chromatography on an EASY-nLC 1200 system coupled to a Q Exactive HF mass spectrometer (Thermo Fisher Scientific). Peptides were separated on a C18-reversed phase column (25 cm long, 75  $\mu\text{m}$  inner diameter) packed in-house with ReproSil-Pur C18-AQ 1.9  $\mu\text{m}$  resin (Dr Maisch). The column was mounted on an Easy Flex Nano Source and temperature controlled by a column oven (Sonation) at 40 °C. A 215-min gradient from 2 to 40% acetonitrile in 0.5% formic acid at a flow of 225 nl/min was used. Spray voltage was set to 2.4 kV. The Q Exactive HF was operated with a TOP20 MS/MS spectra acquisition method per MS full scan. MS scans were conducted with 60,000 at a maximum injection time of 20 ms and MS/MS scans with 15,000 resolution at a maximum injection time of 50 ms. The raw files were processed with MaxQuant version 1.5.2.8 using the LFQ quantification option [71] on unique+razor peptides with at least 2 ratio counts. Carbamidomethylation was set as fixed modification while methionine oxidation and protein N-acetylation were considered as variable modifications. Search results were filtered with a false discovery rate of 0.01. Known contaminants, proteins groups only identified by site, and reverse hits of the MaxQuant results were removed. LFQ intensities can be found in Supplementary data file (GC Proteomics). Differential analysis (log base<sub>2</sub> fold change < 2, adjusted *p* value < 0.05) revealed 64 downregulated proteins in GC CIMP [*n* = 6] vs non-CIMP and non-malignant cell lines [*n* = 8] (Additional files 4, 5).

#### Gene expression and pathway analysis

Differential gene expression of GC CIMP [*n* = 6] vs non-CIMP and non-malignant cell lines [*n* = 8] was performed using Deseq2 (log base<sub>2</sub> fold change  $\leq$  2, adjusted *p* value < 0.05). For the pan-cancer analysis, 22 cancers (TCGA abbreviations: LIHC, HNSC, ESCA, THYM, BLCA, UCEC, ACC, LGG, BRCA, CESC, CHOL, COAD-READ, GBM, KIRP, LUAD, MESO, PAAD, PRAD, SARC, SKCM, TGCT, and UVM) were analyzed for differential gene expression analysis of *CBS* in most “hypermethylated/CIMP” vs. other methylation categories (Wilcoxon rank sum test, *P* < 0.05), as defined by TCGA (The Cancer Genome Atlas) or by Karpinski et al (for UCEC) [52] via MEXPRESS [72]. For UCEC, the TCGA-CIMP sample labels were not publicly available, and therefore, we used CIMP sample labels from Karpinski et al (a pan-cancer study of CIMP). For the GES1 *CBS*-deficient cells (P15) and HFE145 *CBS*-deficient cells (P23), a log base<sub>2</sub> fold change was generated for each clone compared to parental and CRISPR control cells and further filtered with a minimum difference of 1 FPKM (fragments per kilobase of exon model per million reads mapped). Only genes with a linear fold difference (up-regulated [1.5 fold change]) and downregulated [0.6 fold change]) were used for the gene set overlap analysis. For TCGA and CCLE (Cancer Cell Line Encyclopedia) data, genes positively and negatively correlated with *CBS* expression were obtained from cbiportal and filtered using *q*-values < 0.05. MSigDB overlap was computed with hallmark gene sets (v6.2) with *q*-values < 0.01. For the mouse samples, a log base<sub>2</sub> fold change was generated for each gender in the *Tg-hCBS Cbs*<sup>-/-</sup> category compared to *Cbs*<sup>+/+</sup> and further filtered with a minimum difference of 0.5 FPKM. Differential analysis revealed 413 upregulated (1.5 fold change) and 527 downregulated genes (0.6 fold change) common to both genders, which was used for the PANTHER Go-Slim Biological process analysis (FDR < 0.05). For the MSigDB analysis of mouse samples, a log base<sub>2</sub> fold change was generated for each gender in the *Tg-hCBS Cbs*<sup>-/-</sup> category

compared to *Cbs*<sup>+/+</sup> and further filtered with a minimum difference of 1 FPKM. Differential analysis and conversion of mouse to human gene identifiers by the MSigDB tool (v7.2) indicated the following. For upregulated genes (1.5 fold change), 412/1032 converted in female and 555/1249 converted in male whereas for downregulated genes (0.6 fold change), 962/1768 converted in female and 558/1191 converted in male. Overlap with hallmark gene sets (v7.2) with  $q$ -values < 0.01 was performed with the converted IDs.

#### Identification of candidate epimutations in CIMP group

For the discovery GC cell line panel, we overlapped promoter hypermethylated genes ( $n = 595$ ), with those downregulated at RNA ( $n = 1464$  genes) and protein levels ( $n = 64$ ) in CIMP vs. non-CIMP groups using omic platforms as described in the relevant section. For the recurrence analysis of the candidate epimutations in the expanded GC cell line cohort, RPKM, and methylation values were accessed through the Singapore Gastric Cancer Consortium (<http://www.sgcc.sg/research/unique-resources>). For the independent analysis of primary GCs, promoter associated hypermethylated genes in CIMP (as defined by TCGA and clustering of Singapore cohort methylation data [4]) vs. non-CIMP group ( $\beta$ -value difference  $\geq 0.3$ ) were overlapped with downregulated genes in CIMP vs non-CIMP groups ( $q$ -value < 0.05, fold difference  $\leq 0.6$ , median RPKM value in non-CIMP is  $\geq 1$ ). We used a stringent  $\beta$ -value difference of  $\geq 0.3$  since we were concomitantly assessing the effect of hypermethylation on gene expression. TCGA methylation and gene expression data were derived from Infinium HumanMethylation450 beadchip and RNA sequencing platforms respectively. SG cohort methylation and gene expression data were derived from Infinium HumanMethylation27 beadchip and Affymetrix Human Genome U133 Genechips platforms respectively. Except for *FERMT2* in SG cohort, CpG probes pertaining to the other candidate epimutations were represented in both cohorts.

#### RNA extraction, reverse transcription, and quantitative PCR

Total RNA was extracted using the Qiagen RNeasy mini kit according to the manufacturer's instructions. RNA was converted to cDNA using Improm-II<sup>TM</sup> Reverse Transcriptase (Promega). Quantitative PCR was performed in triplicate using Quantifast SYBR Green PCR kit (Qiagen) on an Applied Biosystems HT7900 Real Time PCR System for 40 cycles using the following primers: *CBS* Exon 10/11 F—5' TGCGGAACTA CATGACCAAG 3' , *CBS* Exon 10/11 R—5' TGAGGTCCTCCTCCTTCAGA 3' , *CBS* Exon 16/17 F—5' GCAGATCCAGTACCACAGCA 3' , *CBS* Exon 16/17 R—5' CGAAGGAGAAGTGGGCAAG 3' , *ACTB* F—5' TCCCTGGAGAAGAGCTACG 3' , *ACTB* R—5' GTAGTTTCGTGGATGCCACA 3' . Delta Ct was calculated as a difference in Ct number between experimental and reference gene.

#### Western blotting

Protein lysates were prepared using ice-cold RIPA buffer (Thermo Fisher Scientific, 89900) containing protease inhibitor cocktail (Nacalai Tesque). Protein concentration was determined using Pierce BCA protein assay (Thermo Scientific, 23225). *Laemmli sample buffer* with reducing agent was added to the lysates and boiled at 100 °C for 5 min. Samples were loaded in each well of a 4–15% Mini-Protean TGX gel (Bio-Rad)

and transferred to the PVDF membrane using a semi-dry blotting system (Bio-Rad). Membranes were probed with monoclonal antibodies against CBS (1 in 500 dilution, Abnova H00000875-M01) or ACTB (1 in 2000 dilution, Sigma A5316) in 5% milk-PBST [phosphate-buffered saline with Tween] at 4 °C overnight. Secondary anti-mouse antibody (LNA931, Amersham) was used at a dilution of 1 in 2000 for 1 h at room temperature. Membranes were developed using Amersham ECL Prime Western Blotting Detection Reagent and imaged using Chemidoc Imaging system (Bio-Rad). Bands were quantified using Image Lab software (Bio-Rad).

### **CBS genetic data**

Somatic mutation and copy number data for the stomach adenocarcinoma TCGA 2014 dataset was obtained from cBioportal. FATHMM scores for the single-nucleotide *CBS* somatic variants were derived from COSMIC database. Germline exomes for the stomach TCGA cohort were aligned and variant called as previously described [73]. Briefly, variants with an ExAC MAF < 0.05 were predicted to be deleterious if they were HIGH impact (via Variant Effect Predictor) or MODERATE impact with a CADD score > 25. An additional pathogenicity score for rs117687681 (R369C) was computed using CScape (0.89). *CBS* promoter methylation values for the 6 mutated cases were calculated as an average of 2 TSS probes, cg22633722 and cg09622447.

### **H3K27me3 chromatin immunoprecipitation [ChIP]**

H3K27me3 ChIP was performed by Zymo Research (ab6002, Abcam). After recovery of ChIP and input DNA, whole-genome amplification was performed using the WGA4 kit (Sigma-Aldrich) and BpmI-WGA primers. Amplified DNAs were purified using PCR purification columns (QIAGEN) and digested with BpmI (New England Biolabs) to remove WGA adapters. Thirty nanograms of amplified DNA was used for each sequencing library preparation (New England Biolabs). Libraries were sequenced on HiSeq2500 (Illumina) to an average depth of 20–30 million reads per library. Sequence reads were mapped against the human reference genome (hg19) using Burrows-Wheeler Aligner (BWA-MEM, version 0.7.10). Reads from histone ChIP-seq were trimmed at the first and last 10 bp prior to alignment. Reads with low mapping quality (MAQP < 10) and PCR duplications were removed using Samtools. Histone ChIP-enriched peaks were detected using CCAT. Positive controls for H3K27me3 enrichment were selected from a prior study [74].

### **Transfection experiments**

SNU1750 cells were transfected with 3 individual siRNAs for *CBS* (Integrated DNA Technologies, hs.Ri.CBS.13.1/13.2/13.3) or negative control (Integrated DNA Technologies, 51-01-14-04) at 5 nM using Lipofectamine RNAiMAX (Thermo Fisher Scientific, 13778150). RNA and protein were harvested at 48 h. For the NF- $\kappa$ B reporter assay [75], GES1 and HFE145 *CBS*-deficient cell lines were transfected with 500 ng of 6X-NF- $\kappa$ B-Luciferase reporter or promoter-less pGL3-Basic vector using Lipofectamine 2000 (Thermo Fisher Scientific). Luciferase activity (Promega, E1500) was measured at 48 h post transfection for basal measurements and at 24 h post transfection for the

GYG 4137 experiment. Fold change in luciferase activity was normalized to the viability of the cells using the CellTiter-Glo® Luminescent cell Viability Assay for all samples.

#### **Cell line treatment with drug**

CIMP cell lines were treated with 1 $\mu$ M Azacytidine [Sigma-Aldrich, A3656] or DMSO for 72 h prior to RNA harvest. GES1 and HFE145 *CBS*-deficient cell lines were treated with 0.5 mM of H<sub>2</sub>S donor GYG 4137 (Cayman Chemical, 13345) or DMSO for 24 h.

#### **Immunohistochemistry**

*CBS* monoclonal antibody (Abnova H00000875-M01) was initially tested for immunohistochemistry application in different buffer and antigen retrieval conditions using human liver sections as a positive control (data not shown), after which the following automated method was selected on the Leica Bond III machine. In brief, antigen retrieval was performed with Bond Epitope Solution 2, with primary antibody dilution of 1 in 400 and incubation time of 20 min. Leica Bond Refine kit was used for detection. Imaging was performed on a Nikon Eclipse Ti. The gastric tissue microarray slides were analyzed by a pathologist and cases were scored “positive” if cytoplasmic staining was found in > 5% of gastric epithelial cells and “negative” if staining was entirely absent or present in < 5% of gastric epithelial cells.

#### **Generation of *CBS* CRISPR clones**

GES1 and HFE145 cells were transiently transfected with CRISPR Cas single format vector with a backbone of U6-gRNA/CMV-Cas9-GFP (Sigma-Aldrich) and containing guide RNAs targeting exon 3 of *CBS* gene (HS0000002142 [clone1], HS0000002144 [clone 2], Sigma-Aldrich) or control vector (CRISPR universal negative control 1, Sigma-Aldrich) using Lipofectamine 2000. After 48 h, ~ 960 GFP positive single cells from each transfection were sorted by the Duke-NUS Flow core facility. After ~ 3 weeks, at least ten clones from each vector were cultured and screened and knockouts verified by loss of *CBS* protein using Western blotting. For each cell line model, 2 *CBS*-deficient clones, 1 CRISPR control, and parental cells were utilized for all downstream experiments.

#### **Metabolic analysis of *CBS* CRISPR clones**

GES1 parental, CRISPR control cells, and *CBS*-deficient clone 1 and clone 2 were cultured and harvested together at P16–P19. HFE145 parental, CRISPR control cells, and *CBS*-deficient clone 1 and clone 2 were cultured and harvested together at P23–P24. Each cell line was harvested at least in duplicate. Cells were collected and quenched using 150 mM cold sodium chloride (Merck, Germany) solution. Following centrifugation at 3400g, the cell supernatant was removed. The cell pellets obtained were then processed using a two-phase liquid–liquid extraction protocol based on the modified method of Blish and Dyer. Briefly, polar and lipid metabolites were extracted by sequential addition of methanol (Optima grade, Fisher Scientific, USA), 3.8 mM tricine (Merck) solution and chloroform (gradient grade, Merck) (1:0.5:1 v/v/v, total 2 mL) to the sample. The mixture was vortexed for 1 min following each addition of solvent. The mixture was then centrifuged at 4 °C, 14,000g for 20 min. This resulted in

separation of the sample into two fractions—the top methanol-tricine solution layer contained the polar metabolites while the bottom chloroform layer contained the lipid species. The top polar layer was collected and the remaining mixture was re-extracted using a mixture of methanol and tricine solution (9:10 v/v), followed by centrifugation at 4 °C, 14,000g for another 10 min. The resulting polar layer was combined together with the first polar extract and the bottom chloroform layer containing the lipid metabolites was collected separately. Both layers were stored at – 80 °C before LC-MS analysis. Each polar extract was analyzed in triplicate using an ultra-performance liquid chromatography system (UPLC, Acquity, Waters, USA) in tandem with a mass spectrometer (Q Exactive, Thermo Scientific, USA). A C18 UPLC column (Acquity UPLC HSS T3 column, 2.1 × 100 mm, 1.8 μm, Waters) was used for separation and the mobile phase comprised of two solvents. “A” being water with 0.1% formic acid (VWR, USA) and “B” being methanol with 0.1% formic acid. The UPLC program is as follows: the column was first equilibrated for 0.5 min at 0.1% B. The gradient was then increased from 0.1% B to 50% B over 8 min before being held at 98% B for 3 min. The column was washed for a further 3 min with 98% acetonitrile (Merck) with 0.1% formic acid and finally equilibrated with 0.1% B for 1.5 min. The solvent flow rate was set at 0.4 mL/min; a column temperature of 30 °C was used. The eluent from the UPLC system was directed into the MS. High-resolution mass spectrometry was then performed in both positive and negative electrospray ionization (ESI) modes, with a mass range of 70 to 1050 m/z and a resolution of 70,000. Sheath and auxiliary gas flow was set at 30.0 and 20.0 (arbitrary units) respectively, with a capillary temperature of 400 °C. The spray voltages were 1.25 kV for positive and 1.5 kV negative mode ionization. Mass calibration was performed using standard calibration solution (Thermo Scientific) prior to injection of the samples. A quality control (QC) sample comprising of equal aliquots of each sample was run at regular intervals during the batch LC-MS runs. Lipid extracts were analyzed in triplicate using an ultra-performance liquid chromatography system (UPLC, Acquity, Waters) in tandem with a mass spectrometer (Xevo G2 QTof, Waters). A C18 UPLC column (Acquity UPLC CSH column, 1.0 × 50 mm, 1.7 μm, Waters) was used for separation and the mobile phase comprised of two solvents: “A” comprising of acetonitrile, methanol, and water (2:2:1) with 0.1% acetic acid (Merck) and 0.025% ammonia solution (VWR, USA), and “B” comprising of isopropanol (Fisher Scientific) with 0.1% acetic acid and 0.025% ammonia solution. The UPLC program is as follows: the column was first equilibrated for 1 min at 1% B with a flow rate of 0.1 mL/min. The gradient was then increased from 1% B to 82.5% B over 9 min before B was further increased to 99% for a 5 min wash at a flow rate of 0.15 mL/min. The column was re-equilibrated for 2.2 min at 1% B. Column temperature was maintained at 45 °C and eluent from the LC system was directed into the MS. High-resolution mass spectrometry was then performed in both positive and negative ESI modes with a mass range of 100 to 1800 m/z and a resolution of ≥ 10,000. Cone and desolvation gas flows was set at 40.0 and 600.0 (L/h) respectively, with a desolvation temperature of 600 °C. The ESI capillary voltages were 2.0 kV for positive mode ionization, and 1.0 kV for negative mode ionization. Mass calibration was performed using sodium formate prior to injection of the samples. Similarly, a quality control (QC) sample comprising of equal aliquots of each sample was run at regular intervals during the batch LC-MS runs. The raw LC-MS data obtained from both polar and lipid extracts was processed

using a XCMS-based peak finding algorithm. The QC samples were first used to adjust for instrumental drift and subsequently the experimental data was normalized based on the total ion intensity. Metabolite identities were confirmed based on mass spectral comparison with available metabolite standards. Both univariate (Student's *t* test with Welch correction) and multivariate (Partial Least Squares Discriminant Analysis, PLS-DA) (SIMCA version 13.0.3, Umetrics, Sartorius, Sweden) approaches were used to identify statistically significant differences in metabolite levels between parental and CRISPR control cells vs. *CBS*-deficient clone 1 and clone 2 (univariate—*p* value < 0.05; multivariate—variable importance for projection, VIP > 1).

For the spent media sample preparation, GES1 parental, CRISPR control cells, and *CBS*-deficient clone 1 were cultured and harvested together at P20. RPMI media containing no cells was used as a control. Twenty microliters of 0.5 M dithiothreitol (DTT) (Bio-Rad, CA, USA) was added to 400  $\mu$ L of medium [30, 76]. The mixture was vortexed for 1 min and incubated at 37 °C for 15 min. After incubation, the mixture was centrifuged at 4 °C, 12,000g for 5 min. The treated medium was transferred into a vial and used immediately for homocysteine detection using LC-MS/MS analysis. For the detection of SAM, SAH, and methionine, the spent medium was analyzed separately without any treatment. The treated medium was analyzed in 4 replicates using an ultra-performance liquid chromatography system (UPLC, Acquity, Waters Corp) in tandem with a triple-quadrupole mass spectrometer (Xevo TQ-S, Waters Corp) operating in ESI-positive mode. The source temperature and desolvation temperature were set at 150 °C and 500 °C, respectively. The cone gas flow was 150 l h<sup>-1</sup> and the desolvation gas flow was 700 l h<sup>-1</sup>. The capillary voltage was 2.90 kV. Compound-dependent MS parameters and multiple-reaction monitoring transitions for the analytes are shown in the table below. A C18 UPLC column (Acquity UPLC BEH column, 2.1  $\times$  100 mm, 1.7  $\mu$ m, Waters) was used for chromatographic separation and the mobile phase comprised of two solvents: "A" comprising of water with 0.1% formic acid (VWR, USA) and "B" comprising of methanol (Fisher Scientific, USA) with 0.1% formic acid. The UPLC program is as follows: the column was first equilibrated for 0.5 min at 0.1% B with a flow rate of 0.4 mL/min. The gradient was then increased from 0.1% B to 10% B over 2.5 min before B was further increased to 50% for 3 min at a flow rate of 0.3 mL/min. The column was re-equilibrated at 0.4 mL/min for 1.5 min at 1% B. Column temperature was maintained at 30 °C and eluent from the LC system was directed into the ESI source of the mass spectrometer. The injection volume was 4  $\mu$ L. Chromatographic peak integration was performed with Targetlynx software (version 4.1 SCN810, Waters Corp). Student's *t* test was used to assess statistical significance between control and *CBS*-deficient groups for homocysteine, SAM, SAH, and methionine measurements.

Optimized compound-dependent MS parameters using Xevo TQ-S mass spectrometer

Analyte	Precursor ion mass (m/z)	Fragment ion mass (m/z)	Dwell time (s)	Cone voltage (V)	Collision energy (V)
Homocysteine	136.0	90.1	0.025	15	16
SAM	399.2	136.2	0.025	62	42
SAH	385.1	134.0	0.025	66	33

**Metabolic analysis of CBS CRISPR clones** (Continued)

Analyte	Precursor ion mass (m/z)	Fragment ion mass (m/z)	Dwell time (s)	Cone voltage (V)	Collision energy (V)
Methionine	150.1	104.7	0.025	40	15

**H<sub>2</sub>S synthesis enzyme activity**

The cellular H<sub>2</sub>S production rate was measured following previously described fluorescent methods with minor modifications [77]. Briefly, cultured cells were homogenized in an ice-cold buffer (50 mM Tris-HCl, pH 8.0, 50 mM NaCl) by passing through a 29G-needle syringe 20 times, followed by centrifugation at 14,000g for 15 min at 4 °C. The protein concentration was determined by Nanodrop with correction to remove DNA/RNA contamination. The reaction mixture contained cell extract, 50 mM Tris-HCl (Sigma-Aldrich Corp.), 1 mM L-cysteine (Sigma- Aldrich Corp.), 1 mM homocysteine (Sigma-Aldrich Corp.), 50 μM pyridoxal-5'-phosphate (MP Biomedicals, Santa Ana, CA, USA), and 10 μM H<sub>2</sub>S probe 7-azido-4-methylcoumarin (Santa Cruz Biotechnology, Santa Cruz, CA, USA). Reactions were incubated for 2 h at 37 °C and fluorescence was measured using ex 365 nm, em 450 nm with Varioskan Flash micro-plate reader. A linear standard curve was generated from reactions with sodium hydrosulfide.

**Methylation array profiling and analysis**

GES1 parental, CRISPR control cells (P8, P17, P21, P26, and P31), and CBS-deficient clone 1 (P8, P15, P19, P24, and P29) and clone 2 (P8, P14, P18, P23, and P28) were cultured and harvested together at indicated time points. HFE145 parental, CRISPR control cells (P5 and P23), and CBS-deficient clone 1 (P5 and P23) and clone 2 (P5 and P23) were cultured and harvested together at indicated time points. Genomic DNA was hybridized on Infinium MethylationEPIC Beadchip according to the manufacturer's specifications. Array normalization was performed using the preprocessFunnorm function from the minfi package [10.1093/bioinformatics/btw691]. Probes where the  $\beta$ -value consistently deviated more than 0.2 compared to both the CRISPR negative control and parental cell line of the same passage were considered differentially methylated. Since the global methylation analysis of the CBS-deficient cell lines were across all CpG contexts, we adopted a more stringent  $\beta$ -value difference of  $\geq 0.2$ . Enrichment analysis for PRC2 related-factors was performed using ReMap 2018 v 1.2 [78]. Gastric CIMP- and EBV-CIMP-specific methylation patterns were analyzed using TCGA Infinium HumanMethylation450 beadchip data, comparing samples in either CIMP subgroup to non-CIMP samples in the same dataset. Comparison of GC methylation subtypes (gastric CIMP, EBV-CIMP, non-CIMP) to GIM methylation subtypes ("Methylation high" or CIMP, "Methylation low" or non-CIMP as previously published [15]) were performed at the CGI level. For this analysis, we chose a more relaxed  $\beta$ -value difference  $\geq 0.1$  as we were comparing at the CGI level, which is usually devoid of methylation. In the CpG context analysis, proportion of probes represented the number of hyper- or hypomethylated CpG sites divided by total number of hyper- or hypomethylated CpG sites, respectively in each sample group. Overlap between DNA methylation changes in CBS-deficient cells with DNA methylation changes observed in

TCGA gastric CIMP and EBV-CIMP tumors was performed on CGI level, considering any CGIs that were affected on the probe level in both groups. For the pan-cancer analysis, *CBS* promoter methylation was calculated as an average of 7 CpG probes in the promoter CGI based on TCGA Infinium HumanMethylation450 beadchip data accessed through MEXPRESS [72]. CIMP groups were identified based on TCGA defined clusters (LIHC, ESCA, BLCA, THYM, HNSC) or from Karpinski et al. [52] for UCEC. Differential methylation analysis of CIMP vs non-CIMP groups, and linear regression to gene expression was performed using Wilcoxon rank sum test, and Spearman's correlation analysis, respectively.

### ***Tg-hCBS Cbs*<sup>-/-</sup> mouse work**

*Tg-hCBS Cbs*<sup>-/-</sup> mice were generated as previously described [79], which lack the endogenous mouse *Cbs* gene but have a wild-type human *CBS*-encoding cDNA under the control of the Zinc-inducible metallothionein promoter. *Tg-hCBS Cbs*<sup>-/-</sup> offspring (6 males and 5 females; no zinc supplementation) was obtained from an incross of *Tg-hCBS Cbs*<sup>-/-</sup> × *Tg-hCBS Cbs*<sup>-/-</sup>, and euthanized for tissue collection at different ages (ranging from 2.1–11.1 months). Mouse stomachs were opened along the greater curvature, washed 3× in cold 1× PBS-10% FBS solution, fixed in 4% formaldehyde and embedded in paraffin for hematoxylin and eosin staining and/or RNA extraction (Qiagen RNeasy FFPE Kit). *Cbs*<sup>+/+</sup> mouse stomachs (3 males and 3 females; ranging from 2.8 to 8.1 months) from C57BL/6J mice (genetic background of *Tg-hCBS Cbs*<sup>-/-</sup> mice) were purchased from *InVivos*, Singapore, and processed alongside the *Tg-hCBS Cbs*<sup>-/-</sup> stomachs.

### **Supplementary Information**

The online version contains supplementary material available at <https://doi.org/10.1186/s13059-021-02375-2>.

**Additional file 1.** Supplementary figures.

**Additional file 2: Table S1.** List of hypermethylated gene promoters in the discovery set of CIMP vs non-CIMP cell lines identified by MeDIP-sequencing ( $P < 0.05$ ).

**Additional file 3: Table S2.** List of downregulated genes at RNA level in the discovery set of CIMP vs non-CIMP cell lines ( $P < 0.05$ ).

**Additional file 4: Table S3.** List of downregulated genes at protein level in the discovery set of CIMP vs non-CIMP cell lines ( $P < 0.05$ ).

**Additional file 5: Table S4.** Protein data (LFQ intensities) for the discovery set of CIMP and non-CIMP cell lines.

**Additional file 6: Table S5.** Recurrence analysis of 6 candidate epimutations in expanded GC cell line cohort.

**Additional file 7: Table S6.** *CBS* promoter methylation levels and gene expression in expanded GC cell line cohort.

**Additional file 8: Table S7.** Overlap of genes with promoter hypermethylation ( $\beta$ -value difference  $\geq 0.3$ ), and significant downregulation at the RNA ( $q$ -value  $< 0.05$ ) level in CIMP compared to non-CIMP tumours in TCGA and Singapore cohorts.

**Additional file 9: Table S8.** *CBS* germline and somatic genetic mutations in the gastric TCGA cohort.

**Additional file 10: Table S9.** List of CpG islands comprising 'GIM-GC common CIMP signature'.

**Additional file 11: Table S10.** List of differentially methylated ( $\beta$ -value difference  $\geq 0.2$ ) conserved CpG probes in GES1 *CBS*-deficient clones.

**Additional file 12: Table S11.** List of 76 genes with hypermethylated CpG sites ( $\beta$ -value difference  $\geq 0.2$ ) and downregulation ( $< 0.6$  FD) in both GES1 *CBS*-deficient clones.

**Additional file 13: Table S12.** List of differentially methylated ( $\beta$ -value difference  $\geq 0.2$ ) conserved CpG probes in HFE145 *CBS*-deficient clones.

**Additional file 14: Table S13.** List of 29 genes with hypermethylated CpG sites ( $\beta$ -value difference  $\geq 0.2$ ) and downregulation ( $< 0.6$  FD) in both HFE145 *CBS*-deficient clones.

**Additional file 15: Table S14.** List of CpG islands overlapping hypermethylated CpG sites in *CBS*-deficient cells and GIM-GC common CIMP signature ( $P < 1 \times 10^{-4}$ , bootstrapping using 10000 samples).



**Additional file 16: Table S15.** List of metabolites profiled in GES1 and HFE145 *CBS*-deficient cells.

**Additional file 17: Table S16.** Overlap of genes in GES1 *CBS*-deficient clones with MSigDB hallmark gene sets ( $q$ -value < 0.01).

**Additional file 18: Table S17.** Overlap of genes in HFE145 *CBS*-deficient clones with MSigDB hallmark gene sets ( $q$ -value < 0.01).

**Additional file 19: Table S18.** Overlap of negatively and positively correlated genes to *CBS* in gastric TCGA ( $q$ -value < 0.05) with MSigDB hallmark gene sets ( $q$ -value < 0.01).

**Additional file 20: Table S19.** Overlap of negatively correlated genes to *CBS* in various TCGA cancers ( $q$ -value < 0.05) with MSigDB hallmark gene sets ( $q$ -value < 0.01).

**Additional file 21: Table S20.** Overlap of negatively correlated genes to *CBS* in CCLE ( $q$ -value < 0.05) with MSigDB hallmark gene sets ( $q$ -value < 0.01).

**Additional file 22: Table S21.** Gene expression analysis of *Cbs*<sup>+/+</sup> vs. *Cbs*<sup>-/-</sup> mice.

**Additional file 23: Table S22.** List of upregulated genes and pathways in *CBS*-deficient mouse stomachs (FDR < 0.05).

**Additional file 24: Table S23.** Top 20 pathways in the MSigDB overlap ( $q$ -value < 0.01) of differentially expressed genes in *CBS*-deficient mouse gastric tissue.

**Additional file 25.** Review history.

### Acknowledgements

We thank Dr. Suhail Ahmed Kabeer Rasheed for gifting the NF- $\kappa$ B reporter plasmid, Dr. Kakoli Das for help with imaging, Duke-NUS Genome Biology Facility, and SingHealth Advanced Molecular Pathology Laboratory for providing molecular and immunohistochemistry-related services, respectively. We also thank Prof. Ito and Dr. Matsuo for advice on mouse experiments.

### Review history

The review history is available as Additional file 25.

### Peer review information

Anahita Bishop was the primary editor of this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

### Authors' contributions

N.P. was involved as Lead in Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Project administration, Supervision, Validation, and Writing. H.K.K, A.B, K.L supported in Conceptualization, Data curation, Formal analysis, and Investigation. S.S, S.C, and Z.W supported in Formal analysis, Investigation and Methodology. H.L supported in Investigation and Resources. V.M supported in Data curation, Formal analysis, and Investigation. C.C supported in Formal analysis, Investigation, and Methodology. C.Y.K, O.X, Z.F.A, M.X, and M.H.L supported in Investigation. M.H.L supported in Investigation. A.L.K.T supported in Project administration. J.P supported in Data curation, Formal analysis, and Investigation. A.L.K.T supported in Administration. S.H.W.T supported in Formal analysis, and Investigation. M.L supported in Funding acquisition and Resources. D.K supported in Formal analysis, Investigation, Resources and Supervision. W.K, J.B, S.G.R supported in Resources and Supervision. Y.S.H supported in Formal analysis, Investigation, Resources and Supervision. M.T supported in Investigation, Resources and Supervision. P.T was involved as Lead in Conceptualization, Funding acquisition, Resources, Supervision, and Writing. All author(s) read and approved the final manuscript.

### Funding

This work was supported by Khoo Postdoctoral Fellowship Award (Duke-NUS-KPFA/2016/0012) for N.P., Duke Cancer Institute (DCI) Cancer and Environment Pilot Research Award (ID: RL2016-080) and Duke/Duke-NUS Research Collaboration Pilot Project Award (ID: Duke/Duke-NUS/RECA (Pilot)/2017/0038) for P.T. and M.L., National Medical Research Council grants NMRC/STaR/0026/2015 and OF-LCG18May-0023 for P.T.

### Availability of data and materials

The sequencing data generated from the study can be accessed at NCBI SRA (PRJNA628586) [80]. The methylation array data generated from the study can be accessed at NCBI GEO (GSE149411) [81]. The H3K27me3 ChIP data was accessed from NCBI GEO GSE121140 [82]. DNA methylation and gene expression data for all TCGA datasets were obtained from MEXPRESS [83]. DNA methylation and gene expression data for SG cohort was accessed from NCBI GEO GSE30601 [4] and GSE15460 respectively [84]. DNA methylation data for GIM was accessed from NCBI GEO GSE103186 [15]. *CBS* methylation data for BE was accessed from Krause et al. (Additional file 5, Supplementary Table 4) [18]. Infinium HumanMethylation450 beadchip array data, gene expression, and mutational status of select cancer genes for the GC cell lines were accessed through the Singapore Gastric Cancer Consortium (<http://www.sgcc.sg/research/unique-resources>). Other datasets supporting the conclusions of this article are included within the article as additional files. All other datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

## Declarations

### Ethics approval and consent to participate

Primary GC patient samples for quantitative PCR validation were obtained from the SingHealth tissue repository with approval from the SingHealth Centralised IRB (SingHealth CIRB 2019/2617) and signed patient informed consent, and complied with the Helsinki Declaration. *Tg-hCBS Cbs<sup>-/-</sup>* animal studies were reviewed and approved by the Fox Chase Cancer Center Institutional Animal Care and Review Board.

### Consent for publication

Not applicable.

### Competing interests

The authors declare that they have no competing interests.

### Author details

<sup>1</sup>Programme in Cancer and Stem Cell Biology, Duke-NUS Medical School, 8, College road, Singapore 169857, Singapore. <sup>2</sup>Centre for Computational Biology, Duke-NUS Medical School, Singapore 169857, Singapore. <sup>3</sup>Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 119228, Singapore. <sup>4</sup>Bioprocessing Technology Institute, A\*STAR, 20 Biopolis Way, #06-01 Centros, Singapore 138668, Singapore. <sup>5</sup>Department of Pharmacology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 117600, Singapore. <sup>6</sup>Cancer Biology Program, Fox Chase Cancer Center, Philadelphia, PA, USA. <sup>7</sup>Cancer Science Institute of Singapore, National University of Singapore, Singapore 117599, Singapore. <sup>8</sup>Department of Molecular Genetics and Microbiology, Duke Centre for Virology, Duke University School of Medicine, Durham, NC, USA. <sup>9</sup>Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 117596, Singapore. <sup>10</sup>National University of Singapore (Suzhou) Research Institute, Suzhou 215123, China. <sup>11</sup>Department of Pathology, National University of Singapore, Singapore 119228, Singapore. <sup>12</sup>Genome Institute of Singapore, Singapore 138672, Singapore. <sup>13</sup>SingHealth/Duke-NUS Institute of Precision Medicine, National Heart Centre Singapore, Singapore 169856, Singapore. <sup>14</sup>Singapore Gastric Cancer Consortium, Singapore 119074, Singapore. <sup>15</sup>Department of Physiology, National University of Singapore, Singapore 117593, Singapore.

Received: 28 January 2021 Accepted: 6 May 2021

Published online: 01 June 2021

## References

- Banno K, Kisu I, Yanokura M, Tsuji K, Masuda K, Ueki A, et al. Epimutation and cancer: a new carcinogenic mechanism of Lynch syndrome (Review). *Int J Oncol*. 2012;41(3):793–7. <https://doi.org/10.3892/ijo.2012.1528>.
- Miller BF, Sánchez-Vega F, Elnitski L. The emergence of pan-cancer CIMP and its elusive interpretation. *Biomolecules*. 2016;6(4):45. <https://doi.org/10.3390/biom6040045>.
- Zhao R, Choi BY, Lee M-H, Bode AM, Dong Z. Implications of genetic and epigenetic alterations of CDKN2A (p16<sup>INK4a</sup>) in cancer. *EBioMedicine*. 2016;8:30–9. <https://doi.org/10.1016/j.ebiom.2016.04.017>.
- Zouridis H, Deng N, Ivanova T, Zhu Y, Wong B, Huang D, et al. Methylation subtypes and large-scale epigenetic alterations in gastric cancer. *Sci Transl Med*. 2012;4:156ra140.
- Fennell L, Dumenil T, Wockner L, Hartel G, Nones K, Bond C, et al. Integrative genome-scale DNA methylation analysis of a large and unselected cohort reveals 5 distinct subtypes of colorectal adenocarcinomas. *Cell Mol Gastroenterol Hepatol*. 2019;8(2):269–90. <https://doi.org/10.1016/j.jcmgh.2019.04.002>.
- Sánchez-Vega F, Gotea V, Margolin G, Elnitski L. Pan-cancer stratification of solid human epithelial tumors and cancer cell lines reveals commonalities and tissue-specific features of the CpG island methylator phenotype. *Epigenetics Chromatin*. 2015;8(1):14. <https://doi.org/10.1186/s13072-015-0007-7>.
- Letouzé E, Martinelli C, Loriot C, Burnichon N, Abermil N, Ottolenghi C, et al. SDH mutations establish a hypermethylator phenotype in paraganglioma. *Cancer Cell*. 2013;23(6):739–52. <https://doi.org/10.1016/j.ccr.2013.04.018>.
- Tao Y, Kang B, Petkovich DA, Bhandari YR, In J, Stein-O'Brien G, et al. Aging-like spontaneous epigenetic silencing facilitates wnt activation, stemness, and Braf(V600E)-induced tumorigenesis. *Cancer Cell*. 2019;35(2):315–28 e316. <https://doi.org/10.1016/j.ccell.2019.01.005>.
- Serra RW, Fang M, Park SM, Hutchinson L, Green MR. A KRAS-directed transcriptional silencing pathway that mediates the CpG island methylator phenotype. *Elife*. 2014;3:e02313. <https://doi.org/10.7554/eLife.02313>.
- Cancer Genome Atlas Research N. Comprehensive molecular characterization of gastric adenocarcinoma. *Nature*. 2014; 513(7517):202–9. <https://doi.org/10.1038/nature13480>.
- Matsusaka K, Kaneda A, Nagae G, Ushiku T, Kikuchi Y, Hino R, et al. Classification of Epstein-Barr virus-positive gastric cancers by definition of DNA methylation epigenotypes. *Cancer Res*. 2011;71(23):7187–97. <https://doi.org/10.1158/0008-5472.CAN-11-1349>.
- Matsusaka K, Funata S, Fukuyo M, Seto Y, Aburatani H, Fukayama M, et al. Epstein-Barr virus infection induces genome-wide de novo DNA methylation in non-neoplastic gastric epithelial cells. *J Pathol*. 2017;242(4):391–9. <https://doi.org/10.1002/path.4909>.
- Moarii M, Reyat F, Vert J-P. Integrative DNA methylation and gene expression analysis to assess the universality of the CpG island methylator phenotype. *Human Genomics*. 2015;9(1):26. <https://doi.org/10.1186/s40246-015-0048-9>.
- Zhao H, Li Q, Wang J, Su X, Ng KM, Qiu T, et al. Frequent epigenetic silencing of the folate-metabolising gene cystathionine-beta-synthase in gastrointestinal cancer. *PLoS One*. 2012;7(11):e49683. <https://doi.org/10.1371/journal.pone.0049683>.
- Huang KK, Ramnarayanan K, Zhu F, Srivastava S, Xu C, Tan ALK, et al. Genomic and epigenomic profiling of high-risk intestinal metaplasia reveals molecular determinants of progression to gastric cancer. *Cancer Cell*. 2018;33(1):137–50 e135. <https://doi.org/10.1016/j.ccell.2017.11.018>.

16. Medeiros J-VR, Soares PMG, Brito GADC, Souza MHLPD. Immunohistochemical approach reveals localization of cystathionine- $\gamma$ -lyase and cystathionine- $\beta$ -synthetase in ethanol-induced gastric mucosa damage in mice. *Arq Gastroenterol*. 2013;50:157–60.
17. de Vries AC, van Grieken NCT, Looman CWN, Casparie MK, de Vries E, Meijer GA, et al. Gastric cancer risk in patients with premalignant gastric lesions: a nationwide cohort study in the Netherlands. *Gastroenterology*. 2008;134(4):945–52. <https://doi.org/10.1053/j.gastro.2008.01.071>.
18. Krause L, Nones K, Loffler KA, Nancarrow D, Oey H, Tang YH, et al. Identification of the CIMP-like subtype and aberrant methylation of members of the chromosomal segregation and spindle assembly pathways in esophageal adenocarcinoma. *Carcinogenesis*. 2016;37(4):356–65. <https://doi.org/10.1093/carcin/bgw018>.
19. Choumenkovitch SF, Selhub J, Bagley PJ, Maeda N, Nadeau MR, Smith DE, et al. In the cystathionine beta-synthase knockout mouse, elevations in total plasma homocysteine increase tissue S-adenosylhomocysteine, but responses of S-adenosylmethionine and DNA methylation are tissue specific. *J Nutr*. 2002;132(8):2157–60. <https://doi.org/10.1093/jn/132.8.2157>.
20. Wang X, Zhang S, Zhang J, Lam E, Liu X, Sun J, et al. Annexin A6 is down-regulated through promoter methylation in gastric cancer. *Am J Transl Res*. 2013;5(5):555–62.
21. Cantrell VA, Jessen JR. The planar cell polarity protein Van Gogh-Like 2 regulates tumor cell migration and matrix metalloproteinase-dependent invasion. *Cancer Lett*. 2010;287(1):54–61. <https://doi.org/10.1016/j.canlet.2009.05.041>.
22. Pan K, Liang X-T, Zhang H-K, Zhao J-J, Wang D-D, Li J-J, et al. Characterization of bridging integrator 1 (BIN1) as a potential tumor suppressor and prognostic marker in hepatocellular carcinoma. *Mol Med (Cambridge, Mass)*. 2012;18:507–18.
23. Ward AK, Mellor P, Smith SE, Kendall S, Just NA, Vizeacoumar FS, et al. Epigenetic silencing of CREB3L1 by DNA methylation is associated with high-grade metastatic breast cancers with poor prognosis and is prevalent in triple negative breast cancers. *Breast Cancer Res*. 2016;18(1):12. <https://doi.org/10.1186/s13058-016-0672-x>.
24. Dallol A, Forgacs E, Martinez A, Sekido Y, Walker R, Kishida T, et al. Tumour specific promoter region methylation of the human homologue of the *Drosophila* Roundabout gene DUTT1 (ROBO1) in human cancers. *Oncogene*. 2002;21(19):3020–8. <https://doi.org/10.1038/sj.onc.1205421>.
25. Jiang Z, Liang G, Xiao Y, et al. Targeting the SLIT/ROBO pathway in tumor progression: molecular mechanisms and therapeutic perspectives. *Ther Adv Med Oncol*. 2019;11:1758835919855238. Published 2019 Jun 6. <https://doi.org/10.1177/1758835919855238>.
26. Ulanovskaya OA, Zuhl AM, Cravatt BF. NNMT promotes epigenetic remodeling in cancer by creating a metabolic methylation sink. *Nat Chem Biol*. 2013;9(5):300–6. <https://doi.org/10.1038/nchembio.1204>.
27. Hughey CC, Trefts E, Bracy DP, James FD, Donahue EP, Wasserman DH. Glycine N-methyltransferase deletion in mice diverts carbon flux from gluconeogenesis to pathways that utilize excess methionine cycle intermediates. *J Biol Chem*. 2018;293(30):11944–54. <https://doi.org/10.1074/jbc.RA118.002568>.
28. Varela-Rey M, Martínez-López N, Fernández-Ramos D, Embade N, Calvís DF, Woodhoo A, et al. Fatty liver and fibrosis in glycine N-methyltransferase knockout mice is prevented by nicotinamide. *Hepatology*. 2010;52:105–14. <https://doi.org/10.1002/hep.23639>.
29. Namekata K, Enokido Y, Ishii I, Nagai Y, Harada T, Kimura H. Abnormal lipid metabolism in cystathionine  $\beta$ -synthase-deficient mice, an animal model for hyperhomocysteinemia. *J Biol Chem*. 2004;279:52961–9. <https://doi.org/10.1074/jbc.M406820200>.
30. Hultberg B, Andersson A, Isaksson A. Higher export rate of homocysteine in a human endothelial cell line than in other human cell lines. *Biochim Biophys Acta Mol Cell Res*. 1998;1448(1):61–9. [https://doi.org/10.1016/S0167-4889\(98\)00119-0](https://doi.org/10.1016/S0167-4889(98)00119-0).
31. Zhu H, Blake S, Chan KT, Pearson RB, Kang J. Cystathionine  $\beta$ -synthase in physiology and cancer. *BioMed Res Int*. 2018;2018:11.
32. Paul BD, Snyder SH. H2S signalling through protein sulfhydration and beyond. *Nat Rev Mol Cell Biol*. 2012;13(8):499–507. <https://doi.org/10.1038/nrm3391>.
33. Bourque C, Zhang Y, Fu M, Racine M, Greasley A, Pei Y, et al. H2S protects lipopolysaccharide-induced inflammation by blocking NF $\kappa$ B transactivation in endothelial cells. *Toxicol Appl Pharmacol*. 2018;338:20–9. <https://doi.org/10.1016/j.taap.2017.11.004>.
34. Robert K, Nehme J, Bourdon E, Pivert G, Friguet B, Delcayre C, et al. Cystathionine beta synthase deficiency promotes oxidative stress, fibrosis, and steatosis in mice liver. *Gastroenterology*. 2005;128(5):1405–15. <https://doi.org/10.1053/j.gastro.2005.02.034>.
35. Guo F-F, Yu T-C, Hong J, Fang J-Y. Emerging roles of hydrogen sulfide in inflammatory and neoplastic colonic diseases. *Frontiers in physiology*. 2016;7:156. <https://doi.org/10.3389/fphys.2016.00156>.
36. Ghandi M, Huang FW, Jané-Valbuena J, Kryukov GV, Lo CC, McDonald ER, et al. Next-generation characterization of the Cancer Cell Line Encyclopedia. *Nature*. 2019;569(7757):503–8. <https://doi.org/10.1038/s41586-019-1186-3>.
37. Gupta S, Kühnisch J, Mustafa A, Lhotak S, Schlachterman A, Slifker MJ, et al. Mouse models of cystathionine beta-synthase deficiency reveal significant threshold effects of hyperhomocysteinemia. *FASEB J*. 2009;23(3):883–93. <https://doi.org/10.1096/fj.08-120584>.
38. Cal-Kayitmazbatir S, Kulkoyluoglu-Cotul E, Growe J, Selby CP, Rhoades SD, Malik D, Oner H, Asimgil H, Francey LJ, Sancar A, et al. CRY1-CBS binding regulates circadian clock function and metabolism. 2020:2020.2001.2009.898866.
39. Issa JP. CpG island methylator phenotype in cancer. *Nat Rev Cancer*. 2004;4(12):988–93. <https://doi.org/10.1038/nrc1507>.
40. Stylianou E. Epigenetics of chronic inflammatory diseases. *J Inflamm Res*. 2018;12:1–14. <https://doi.org/10.2147/JIRS.129027>.
41. Yamashita S, Nanjo S, Rehnberg E, Iida N, Takeshima H, Ando T, et al. Distinct DNA methylation targets by aging and chronic inflammation: a pilot study using gastric mucosa infected with *Helicobacter pylori*. *Clin Epigenetics*. 2019;11(1):191. <https://doi.org/10.1186/s13148-019-0789-8>.
42. Merry CR, Forrest ME, Sabers JN, Beard L, Gao X-H, Hatzoglou M, et al. DNMT1-associated long non-coding RNAs regulate global gene expression and DNA methylation in colon cancer. *Hum Mol Genet*. 2015;24(21):6240–53. <https://doi.org/10.1093/hmg/ddv343>.
43. Cao S, Zhu X, Zhang C, Qian H, Schuttler H-B, Gong J, Xu Y. Competition between DNA methylation, nucleotide synthesis, and antioxidation in cancer versus normal tissues. 2017, 77:4185–4195. <https://doi.org/10.1158/0008-5472.CA.N-17-0262>.
44. Tang B, Mustafa A, Gupta S, Melnyk S, James SJ, Kruger WD. Methionine-deficient diet induces post-transcriptional downregulation of cystathionine  $\beta$ -synthase. *Nutrition*. 2010;26(11-12):1170–5. <https://doi.org/10.1016/j.nut.2009.10.006>.

45. Zhu H, Blake S, Chan KT, Pearson RB, Kang J. Cystathionine  $\beta$ -synthase in physiology and cancer. *BioMed Res Int.* 2018; 2018:3205125.
46. Niwa T, Tsukamoto T, Toyoda T, Mori A, Tanaka H, Maekita T, et al. Inflammatory processes triggered by *Helicobacter pylori* infection cause aberrant DNA methylation in gastric epithelial cells. *Cancer Res.* 2010;70(4):1430–40. <https://doi.org/10.1158/0008-5472.CAN-09-2755>.
47. Erchegaray J-P, Mostoslavsky R. Interplay between metabolism and epigenetics: a nuclear adaptation to environmental changes. *Mol Cell.* 2016;62(5):695–711. <https://doi.org/10.1016/j.molcel.2016.05.029>.
48. Yang R, Qu C, Zhou Y, Konkel JE, Shi S, Liu Y, et al. Hydrogen sulfide promotes Tet1- and Tet2-mediated Foxp3 demethylation to drive regulatory T cell differentiation and maintain immune homeostasis. *Immunity.* 2015;43(2):251–63. <https://doi.org/10.1016/j.immuni.2015.07.017>.
49. Liu Y, Mayo MW, Nagji AS, Smith PW, Ramsey CS, Li D, et al. Phosphorylation of RelA/p65 promotes DNMT-1 recruitment to chromatin and represses transcription of the tumor metastasis suppressor gene BRMS1. *Oncogene.* 2012; 31(9):1143–54. <https://doi.org/10.1038/onc.2011.308>.
50. O'Hagan HM, Wang W, Sen S, Destefano Shields C, Lee SS, Zhang YW, et al. Oxidative damage targets complexes containing DNA methyltransferases, SIRT1, and polycomb members to promoter CpG Islands. *Cancer Cell.* 2011;20(5): 606–19. <https://doi.org/10.1016/j.ccr.2011.09.012>.
51. García-Giménez JL, Pallardó FV. Maintenance of glutathione levels and its importance in epigenetic regulation. *Front Pharmacol.* 2014;5:88.
52. Karpinski P, Pesz K, Sasiadek MM. Pan-cancer analysis reveals presence of pronounced DNA methylation drift in CpG island methylator phenotype clusters. *Epigenomics.* 2017;9(11):1341–52. <https://doi.org/10.2217/epi-2017-0070>.
53. Luebeck EG, Curtius K, Hazelton WD, Maden S, Yu M, Thota PN, et al. Identification of a key role of widespread epigenetic drift in Barrett's esophagus and esophageal adenocarcinoma. *Clin Epigenetics.* 2017;9(1):113. <https://doi.org/10.1186/s13148-017-0409-4>.
54. Luebeck GE, Hazelton WD, Curtius K, Maden SK, Yu M, Carter KT, et al. Implications of epigenetic drift in colorectal neoplasia. *Cancer Res.* 2019;79(3):495–504. <https://doi.org/10.1158/0008-5472.CAN-18-1682>.
55. Shaposhnikov M, Proshkina E, Koval L, Zemskaya N, Zhavoronkov A, Moskalev A. Overexpression of CBS and CSE genes affects lifespan, stress resistance and locomotor activity in *Drosophila melanogaster*. *Aging.* 2018;10(11):3260–72. <https://doi.org/10.18632/aging.101630>.
56. Singh SB, Lin HC. Hydrogen sulfide in physiology and diseases of the digestive tract. *Microorganisms.* 2015;3(4):866–89. <https://doi.org/10.3390/microorganisms3040866>.
57. Cao, Xu et al. "A Review of Hydrogen Sulfide Synthesis, Metabolism, and Measurement: Is Modulation of Hydrogen Sulfide a Novel Therapeutic for Cancer?." *Antioxidants & redox signaling.* 2019;31(1):1-38. <https://doi.org/10.1089/ars.2017.7.7058>.
58. Zhao K, Ju Y, Li S, Altaany Z, Wang R, Yang G. S-sulfhydration of MEK1 leads to PARP-1 activation and DNA damage repair. *EMBO Rep.* 2014;15(7):792–800. <https://doi.org/10.1002/embr.201338213>.
59. Zhang D, Du J, Tang C, Huang Y, Jin H. H<sub>2</sub>S-induced sulfhydration: biological function and detection methodology. *Front Pharmacol.* 2017;8:608.
60. Behera J, Kelly KE, Voor MJ, Metreveli N, Tyagi SC, Tyagi N. Hydrogen sulfide promotes bone homeostasis by balancing inflammatory cytokine signaling in CBS-deficient mice through an epigenetic mechanism. *Sci Rep.* 2018;8(1):15226. <https://doi.org/10.1038/s41598-018-33149-9>.
61. Magierowski M, Wierdak M, Magierowska K, Janmaat V, Hubalewska-Mazgaj M, Chmura A, et al. Sa1099 - hydrogen sulfide (H<sub>2</sub>S) prevents development of Barrett's esophagus metaplasia. *Gastroenterology.* 2018;154:S–240.
62. Wallace JL, Dickey M, McKnight W, Martin GR. Hydrogen sulfide enhances ulcer healing in rats. *Faseb j.* 2007;21(14):4070–6. <https://doi.org/10.1096/fj.07-8669com>.
63. Bazhanov N, Escaffre O, Freiberg AN, Garofalo RP, Casola A. Broad-range antiviral activity of hydrogen sulfide against highly pathogenic RNA viruses. *Sci Rep.* 2017;7(1):41029. <https://doi.org/10.1038/srep41029>.
64. Wallace JL, Vaughan D, Dickey M, MacNaughton WK, de Nucci G. Hydrogen sulfide-releasing therapeutics: translation to the clinic. *Antioxid Redox Signal.* 2018;28(16):1533–40. <https://doi.org/10.1089/ars.2017.7068>.
65. Ku JL, Kim KH, Choi JS, Kim SH, Shin YK, Chang HJ, et al. Establishment and characterization of six human gastric carcinoma cell lines, including one naturally infected with Epstein-Barr virus. *Cell Oncol (Dordr).* 2012;35(2):127–36. <https://doi.org/10.1007/s13402-012-0073-9>.
66. Oh ST, Seo JS, Moon UY, Kang KH, Shin DJ, Yoon SK, et al. A naturally derived gastric cancer cell line shows latency I Epstein-Barr virus infection closely resembling EBV-associated gastric cancer. *Virology.* 2004;320(2):330–6. <https://doi.org/10.1016/j.virol.2003.12.005>.
67. Kim DN, Seo MK, Choi H, Kim SY, Shin HJ, Yoon AR, et al. Characterization of naturally Epstein-Barr virus-infected gastric carcinoma cell line YCCEL1. *J Gen Virol.* 2013;94(3):497–506. <https://doi.org/10.1099/vir.0.045237-0>.
68. Yamamoto H, Itoh F, Fukushima H, Hinoda Y, Imai K. Overexpression of cyclooxygenase-2 protein is less frequent in gastric cancers with microsatellite instability. *Int J Cancer.* 1999;84(4):400–3. [https://doi.org/10.1002/\(SICI\)1097-0215\(19990820\)84:4<400::AID-IJC12>3.0.CO;2-5](https://doi.org/10.1002/(SICI)1097-0215(19990820)84:4<400::AID-IJC12>3.0.CO;2-5).
69. Cristescu R, Lee J, Nebozhyn M, Kim KM, Ting JC, Wong SS, et al. Molecular analysis of gastric cancer identifies subtypes associated with distinct clinical outcomes. *Nat Med.* 2015;21(5):449–56. <https://doi.org/10.1038/nm.3850>.
70. Radford EJ, Ito M, Shi H, Corish JA, Yamazawa K, Isganaitis E, et al. In utero effects. In utero undernourishment perturbs the adult sperm methylome and intergenerational metabolism. *Science.* 2014;345:1255903.
71. Cox J, Hein MY, Lubner CA, Paron I, Nagaraj N, Mann M. Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol Cell Proteomic.* 2014;13(9):2513–26. <https://doi.org/10.1074/mcp.M113.031591>.
72. Koch A, Jeschke J, Van Criekinge W, van Engeland M, De Meyer T. MEXPRESS update 2019. *Nucleic Acids Res.* 2019; 47(W1):W561–5. <https://doi.org/10.1093/nar/gkz445>.
73. Pitt JJ, Riester M, Zheng Y, Yoshimatsu TF, Sanni A, Oluwasola O, et al. Characterization of Nigerian breast cancer reveals prevalent homologous recombination deficiency and aggressive molecular features. *Nat Commun.* 2018;9(1):4181. <https://doi.org/10.1038/s41467-018-06616-0>.

74. Gao F, Ji G, Gao Z, Han X, Ye M, Yuan Z, et al. Direct ChIP-bisulfite sequencing reveals a role of H3K27me3 mediating aberrant hypermethylation of promoter CpG islands in cancer cells. *Genomics*. 2014;103(2-3):204–10. <https://doi.org/10.1016/j.ygeno.2013.12.006>.
75. Rasheed SAK, Leong HS, Lakshmanan M, Raju A, Dadlani D, Chong FT, et al. GNA13 expression promotes drug resistance and tumor-initiating phenotypes in squamous cell cancers. *Oncogene*. 2018;37(10):1340–53. <https://doi.org/10.1038/s41388-017-0038-6>.
76. Persichilli S, Gervasoni J, Iavarone F, Zuppi C, Zappacosta B. A simplified method for the determination of total homocysteine in plasma by electrospray tandem mass spectrometry. *J Sep Sci*. 2010;33(20):3119–24. <https://doi.org/10.1002/jssc.201000399>.
77. Gersztenkorn D, Coletta C, Zhu S, Ha Y, Liu H, Tie H, et al. Hydrogen sulfide contributes to retinal neovascularization in ischemia-induced retinopathy. *Invest Ophthalmol Vis Sci*. 2016;57(7):3002–9. <https://doi.org/10.1167/iov.15-18555>.
78. Chèneby J, Gheorghe M, Artufel M, Mathelier A, Ballester B. ReMap 2018: an updated atlas of regulatory regions from an integrative analysis of DNA-binding ChIP-seq experiments. *Nucleic Acids Res*. 2017;46:D267–75.
79. Wang L, Jhee K-H, Hua X, DiBello PM, Jacobsen DW, Kruger WD. Modulation of cystathionine  $\beta$ -synthase level regulates total serum homocysteine in mice. *Circ Res*. 2004;94(10):1318–24. <https://doi.org/10.1161/01.RES.0000129182.46440.4a>
80. Nisha Padmanabhan, Huang Kie Kyon, Arnoud Boot, Kevin Lim, Supriya Srivastava, Shuwen Chen, Zhiyuan Wu, OK Hyung-Lee, Vineeth T. Mukundan, Charlene Chan, Yarn Kit Chan, Ong Xuewen, Jason J. Pitt, Zul Fazreen Adam Isa, Man.jie Xing, Ming Hui Le, Angie Lay Keng Tan, Shamaine Ho Wei Ting, Micah A. Luftig, Dennis Kappei, Warren D. Kruger, Jinsong Bian, Ying Swan Ho, Ming Teh, Steve George Rozen and Patrick Tan. Highly recurrent CBS epimutations in gastric cancer CpG island methylator phenotypes and inflammation. 2021. sequencing data: <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA628586>.
81. Nisha Padmanabhan, Huang Kie Kyon, Arnoud Boot, Kevin Lim, Supriya Srivastava, Shuwen Chen, Zhiyuan Wu, Hyung-OK Lee, Vineeth T. Mukundan, Charlene Chan, Yarn Kit Chan, Ong Xuewen, Jason J. Pitt, Zul Fazreen Adam Isa, Man.jie Xing, Ming Hui Le, Angie Lay Keng Tan, Shamaine Ho Wei Ting, Micah A. Luftig, Dennis Kappei, Warren D. Kruger, Jinsong Bian, Ying Swan Ho, Ming Teh, Steve George Rozen and Patrick Tan. Highly recurrent CBS epimutations in gastric cancer CpG island methylator phenotypes and inflammation. 2021. Methylation array data: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE149411>.
82. Xing M, Ooi WF, Tan J, Qamra A, Lee PH, Li Z, et al. Genomic and epigenomic EBF1 alterations modulate TERT expression in gastric cancer. *J Clin Invest*. 2020;130(6):3005–20. <https://doi.org/10.1172/JCI126726>.
83. Koch A, De Meyer T, Jeschke J, Van Criekinge W. MEXPRESS: visualizing expression, DNA methylation and clinical TCGA data. *BMC Genomics*. 2015;16:636.
84. Ooi CH, Ivanova T, Wu J, Lee M, Tan IB, Tao J, et al. Oncogenic pathway combinations predict clinical prognosis in gastric cancer. *PLoS Genet*. 2009;5(10):e1000676. <https://doi.org/10.1371/journal.pgen.1000676>.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

