Journal of Inorganic Biochemistry 251 (2024) 112422

Contents lists available at ScienceDirect

Journal of Inorganic Biochemistry

journal homepage: www.elsevier.com/locate/jinorgbio

Molecular analysis of the human cytoglobin mRNA isoforms

Elena Porto [a], Paraskevi Loula [a], Susanne Strand [b], Thomas Hankeln [a,*]

[a] Institute of Organismic and Molecular Evolution, Molecular Genetics & Genome Analysis Group, Johannes Gutenberg University Mainz, J. J. Becher-Weg 30A, D-55128 Mainz, Germany

[b] Department of Internal Medicine I, Molecular Hepatology, University Medical Center, Johannes Gutenberg University Mainz, Obere Zahlbacher Strasse 63, 55131 Mainz, Germany

ARTICLE INFO

Keywords:
Oxygen-binding
Heme
Alternative splicing

ABSTRACT

Multiple functions have been proposed for the ubiquitously expressed vertebrate globin cytoglobin (Cygb), including nitric oxide (NO) metabolism, lipid peroxidation/signalling, superoxide dismutase activity, reactive oxygen/nitrogen species (RONS) scavenging, regulation of blood pressure, antifibrosis, and both tumour suppressor and oncogenic effects. Since alternative splicing can expand the biological roles of a gene, we investigated whether this mechanism contributes to the functional diversity of *Cygb*. By mining of cDNA data and molecular analysis, we identified five alternative mRNA isoforms for the human *CYGB* gene (V-1 to V-5). Comprehensive RNA-seq analyses of public datasets from human tissues and cells confirmed that the canonical *CYGB* V-1 isoform is the primary *CYGB* transcript in the majority of analysed datasets. Interestingly, we revealed that isoform V-3 represented the predominant *CYGB* variant in hepatoblastoma (HB) cell lines and in the majority of analysed normal and HB liver tissues. *CYGB* V-3 mRNA is transcribed from an alternate upstream promoter and hypothetically encodes a N-terminally truncated CYGB protein, which is not recognized by some antibodies used in published studies. Little to no transcriptional evidence was found for the other *CYGB* isoforms. Comparative transcriptomics and flow cytometry on $CYGB^{+/+}$ and gene-edited $CYGB^{-/-}$ HepG2 HB cells did not unveil a knockout phenotype and, thus, a potential function for *CYGB* V-3. Our study reveals that the *CYGB* gene is transcriptionally more complex than previously described as it expresses alternative mRNA isoforms of unknown function. Additional experimental data are needed to clarify the biological meaning of those alternative *CYGB* transcripts.

1. Introduction

Cytoglobin (Cygb) [1,2] is a highly conserved member of the vertebrate globin superfamily, which includes respiratory metallo-proteins such as the well-studied hemoglobin [3], expressed in erythrocytes, muscle-cell myoglobin [4], nervous system-expressed neuroglobin [5], and the cilia-associated androglobin [6,7]. In contrast to those globins, Cygb is not primarily detected in a few distinct tissues or cell types, but rather is expressed in a large range of tissues [1,8] and cells [9–17]. Originally identified in the cytoplasm of rat hepatic stellate cells (HSCs) [2] and other fibroblasts and fibroblast-like cell types [10,18], Cygb was also reported to be endogenously expressed in distinct neurons

Abbreviations: 5'RACE, rapid amplification of cDNA ends; aa, amino acids; AS, alternative splicing; AUAP, abridged universal amplification primer; CAGE, cap analysis of gene expression; CCNB1, regulators of the G2/M checkpoint cyclin B1 gene; CDK1, cyclin dependent kinase 1 gene; CDS, coding sequence; Cygb, cytoglobin; CYGB S-S monomer, CYGB monomer with an intramolecular disulfide bridge; $CYGB^{-/-}$, *CYGB* knockout; $CYGB^{+/+}$, CYGB wildtype; Cys, cysteine; EST, expressed sequence tag; FDR, false discovery rate; GAPDH, glyceraldehyde 3-phosphate dehydrogenase; GSP1, *CYGB* specific primer 1; GSP2, *CYGB* gene specific primer 2; GSP3, *CYGB* gene specific primer 3; HB, hepatoblastoma; His, histidine; HMEC, human primary mammary epithelial cells; HSCs, hepatic stellate cells; NC, negative control; NHEK, primary normal human epidermal keratinocytes; NiR, nitrite reductase; NO, nitric oxide; ORF, open reading frame; Phe, phenylalanine; PLK1, polo-like kinase 1 gene; Pol II, RNA polymerase II; PRCD, photoreceptor disc component gene; qRT-PCR, quantitative RT-PCR; RNA-PET, paired-end diTag sequencing of RNA; RONS, oxygen/nitrogen species; RPKM, reads per kilobase per million mapped reads; RT-PCR, reverse transcriptase-polymerase chain reaction; SNHG16, small nucleolar RNA host gene 16; T-C1, HB, subtype C1; T-C2, HB, subtype C2; TOP2A, DNA topoisomerase IIα; TSS, transcription starting site; V-1, *CYGB* isoform V-1; V-2, *CYGB* isoform V-2; V-3, *CYGB* isoform V-3; V-3r, *CYGB* isoform V-3r (5'RACE-defined); V-4, *CYGB* isoform V-4; V-5, *CYGB* isoform V-5; VSMCs, vascular smooth muscle cells; WB, western blot.

* Corresponding author.
E-mail address: hankeln@uni-mainz.de (T. Hankeln).

https://doi.org/10.1016/j.jinorgbio.2023.112422
Received 28 April 2023; Received in revised form 26 September 2023; Accepted 29 October 2023
Available online 15 November 2023
0162-0134/© 2023 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

(cytoplasm and nucleus) [9,11], in melanocytes [15], myoblasts [19], vascular smooth muscle cells [20–24], adipocytes [17], and, in contrast to previous findings, in the nucleus and/or cytoplasm of some epithelial cells [9,13,14,25]. Multiple functions of Cygb were proposed and studied *in vitro*: NO dioxygenase [24,26–31], nitrite reductase [20,23,32], superoxide dismutase [33], and lipid peroxidation [34,35] activities, modulation of lipid signalling [36], cytoprotective role via reactive oxygen/nitrogen species (RONS) scavenging [15,37–39], and participation in collagen maturation [10,11]. Few studies investigated the *in vivo* role of Cygb. In a skeletal muscle-specific Cygb knockout mouse model, Cygb was shown to modulate myogenic progenitor viability [37], while studies in a full genetic Cygb knockout model ($Cygb^{-/-}$) demonstrated that, through its role in the NO metabolism of VSMCs, Cygb controls blood pressure and vascular tone [23], regulates neointima formation, and inhibits apoptosis after injury [21]. In a $Cygb^{-/-}$ mouse model of cholestatic liver disease, the lack of Cygb in HSCs was shown to enhance fibrosis, hepatocyte damage, and liver inflammation, possibly by deregulating NO metabolism [40]. In line with these data, a recent study in $Cygb^{-/-}$ mice showed that the NO overexpressed as a consequence of *Cygb* depletion in HSCs diffuses to the surrounding hepatocytes, causing mitochondrial dysfunction and ROS-induced liver damage [41]. Moreover, N,N-deethylnitrosamine-treated $Cygb^{-/-}$ mice showed increased liver and lung tumorigenesis [42], a $Cygb^{-/-}$ mouse model of non-alcoholic steatohepatitis developed liver cancer [43], and aged $Cygb^{-/-}$ mice showed various organ abnormalities [44] compared to their wildtype controls, thus arguing for a tumour suppressor role of Cygb. In line with this, a $Cygb^{-/-}$ mouse model study suggested that Cygb inhibits the development of already existing microadenomas in colon [45]. However, interpretation of Cygb's role in cancer is complex, since it was proposed to function as an oncogene in some cancer types because of its upregulation in hypoxic tumours [14,46,47].

The human *CYGB* gene located at chromosome position 17q25 is composed of 4 exons and 3 introns [1]. Its promoter presents a 1.4 kb CpG island with multiple GC boxes/Sp1 binding sites, but no TATA-box, thus explaining the DNA methylation-dependent epigenetic control of its expression [48]. In fact, *CYGB* silencing via hypermethylation of its promoter was observed in various cancer types [15,49–52]. Multiple hypoxia-responsive elements were found in this promoter [48], explaining *CYGB* upregulation in low-oxygen conditions [53]. The human CYGB is a 21 kDa, 190 amino acid (aa) hexacoordinated protein that contains a consensual globin domain with conserved "proximal" and "distal" histidine residues (His113 and His81, respectively) [1]. Characteristic N- and C-terminal extensions distinguish CYGB from the other vertebrate globins [1], probably contributing to the binding of CYGB to ligands [1,54] and participating in the structural stability of the protein [55]. Two cysteine residues, Cys38 and Cys83, are probably important for the modulation of CYGB functions: under oxidative *in vitro* conditions, these residues contribute to the generation of stable homodimers via the establishment of intermolecular disulfide bridges [56,57], and/or to formation of an intramolecular disulfide bridge in the CYGB monomer (CYGB S-S monomer) [58]. The intramolecular disulfide bridge induces a pentacoordinated-like state in the CYGB monomer that facilitates its binding to exogenous ligands (e.g. lipids and O_2) [34,59–61] and enhances peroxidase [34,35] and NiR activities [32]. Therefore, the cellular redox state significantly influences the structure of CYGB, possibly enabling different CYGB functions in distinct tissues, cells, and oxidative environments.

The process of alternative splicing (AS) is a conserved mechanism by which different mature mRNAs are generated from a single gene, thereby, in generating phenotypic complexity in higher organisms. Due to AS, up to 95% of the human multi-exon genes code for more than one mRNA isoform [62], and RNA-seq studies indicated that 86% of human genes have a minor isoform frequency of at least 15% [63]. Accounting for the complex biochemistry of Cygb and its widespread expression in tissues and cell types, we asked whether AS could contribute to the diversity of functions attributed to this globin. Here we report five putative transcript isoforms, four of them novel, for the human *CYGB* gene (*CYGB* V-1 to *CYGB* V-5) via *in silico* analyses and study their expression in human tissues.

2. Materials and methods

2.1. Data mining in expressed sequence tag (EST) databases

Eleven mRNA variants are reported to be associated with the *CYGB* gene locus in the AceView database (https://www.ncbi.nlm.nih.gov/IEB/Research/Acembly /, last accession October 2022). These mRNA variants are supported by a total of 169 GenBank accessions; the nucleotide sequence of each of these evidences, if available, was downloaded and manually analysed. 66 EST entries were discarded because of mapping to the anti-sense strand of the *CYGB* gene locus: 64 on the photoreceptor disc component (*PRCD*) gene locus and 2 on intronic regions, indicating that they were not originating from a bona fide *CYGB* mRNA molecule. Furthermore, 8 entries had no associated sequence and 3 were not cDNA evidences. To mine data from the UCSC genome browser (https://genome.ucsc.edu/), the EST tracks (human and human spliced) were visualized and all entries mapping correctly to the human *CYGB* gene locus or 20 kb up- or downstream of this region were downloaded. Accordingly, 117 EST entries (63 non-spliced and 54 spliced) were filtered and, among these, all AceView database evidences were included. To identify putative *CYGB* mRNA variants, the collection of ESTs was searched for sequences mapping at least partially to one of the annotated *CYGB* exons. 86 entries were identified, 48 of which represented spliced ESTs. Consensus sequences for putative *CYGB* mRNA variants were generated and used to mine 4 additional cDNA evidences from the EST databases. All EST entries used in the subsequent analysis are listed in Supplementary Table 1.

2.2. Open reading frame (ORF) prediction

CYGB mRNA isoforms identified by EST and 5′ rapid amplification of cDNA ends (5'RACE) analyses (see below) were searched for putative coding regions by examining their nucleotide sequences for ORFs. The ORFfinder tool (https://www.ncbi.nlm.nih.gov/orffinder/) [64] was used with default parameters: minimal ORF length = 75 nt, genetic code = standard, and ORF starting codon = "ATG". Only ORFs predicted on the "+" strand of the investigated sequences were retained for further analyses because the input sequences originated from mRNA molecules.

2.3. Protein structure analysis

Three predicted protein structures are reported for CYGB in the AlphaFold protein structure database (https://alphafold.com/) [65,66]: Q8WWM9, K7EIM9, and K7EMC7, corresponding to the canonical, the truncated, and the V-4 protein isoforms, respectively. The structures were downloaded as .pdb files and compared using the pairwise structure alignment tool of RCSB Protein Data Bank (https://www.rcsb.org/alignment) [67] with default parameters.

2.4. Transcriptome analysis of public RNA-seq datasets and estimation of CYGB mRNA isoform proportions

RNA-seq datasets from tumour and non-tumourous tissues from 24 hepatoblastoma (HB) patients were obtained from the PRJNA413799 study. Among the HB samples, $n = 18$ were classified as subtype C1 (T-C1), a standard-risk HB subtype with a fetal phenotype and low mitotic index, and $n = 6$ as subtype C2 (T-C2), a more immature phenotype with advanced tumour stage, higher cellular proliferation, and worse survival rate [68–71]. Raw data were downloaded from the EMBL-EBI ENA database (www.ebi.ac.uk/ena), the quality of the reads from each dataset was assessed with the FastQC tool (version 0.11.2, http://www.

bioinformatics.babraham.ac.uk/projects/fastqc/), and raw read pre-processing was performed using the FASTX-Toolkit (version 0.0.14, http://hannonlab.cshl.edu/fastx_toolkit/). The pre-processed reads were mapped against the human reference genome (version GRCh37/hg19) with the RNA-seq analysis tool of the CLC Genomics Workbench (version 6.5.1, https://www.qiagenbioinformatics.com/), allowing one single match for a read and a maximum of 2 base mismatches per mapping, while minimum length and minimum similarity fractions were both set to 0.95. The expression level of each gene was calculated and expressed as an RPKM (reads per kilobase permillion mapped reads) value [72] using the RNA-Seq algorithm. The accession number, the sample description, the statistics for the pre-processing and mapping steps, as well as the RPKM values for *CYGB* and the long non-coding RNA gene *SNHG16* are reported in Supplementary File 1. Differential gene expression analysis was performed with the DESeq2 Bioconductor package (version 1.16.1, https://bioconductor.org/packages/release/bioc/html/DESeq2.html) [73] using default parameters and the calculated *p*-values were adjusted for multiple testing using the false discovery rate (FDR) correction. RNA-seq analyses of other human normal tissues, primary cells, and cell lines were performed as described for the PRJNA413799 study. For each sample expressing *CYGB* (*CYGB* RPKM>0.5) the reads mapping to each of the exons, introns, and exon-exon junction regions of *CYGB* were automatically counted and the proportion of each *CYGB* mRNA isoform was estimated based on the isoform-specific splicing events (see Supplementary methods).

2.5. Promoter prediction

The Gene2Promoter tool from Genomatix Genome Analyzer (version v3.40820) was used to identify putative promoters in the *CYGB* locus and in the 20 kb upstream and downstream of this region using as references the human genome assembly GRCh37/hg19 and the Genomatix ElDorado annotation (version 12–2013), which contains, among others, promoter predictions from the PromoterInspector tool of Genomatix. The identified putative promoters were searched for the presence of TATA boxes with the TSSW online tool from Softberry (http://www.softberry.com) [74] and with the online tool GPMiner (http://GPMiner.mbc.nctu.edu.tw/) [75] using default parameters. Similarly, GPMiner was used to identify CCAAT and GC boxes in the putative promoter sequences.

2.6. Evolutionary conservation analysis of CYGB transcript isoforms

The BLASTN algorithm was used, via its dedicated online tool (https://blast.ncbi.nlm.nih.gov/Blast.cgi), to search the NCBI nucleotide database (last accession November 2018) for non-human sequences putatively homologous to the EST- and 5'RACE-defined *CYGB* mRNA isoforms. Default parameters were applied to the algorithm with minor alterations: in the aim of detecting a maximum of sequence hits, the search was optimized to identify up to 20,000 "somewhat similar" sequences. The obtained hits were visually analysed to identify sequences putatively homologous to the *CYGB* isoforms.

2.7. Cell culture

HepG2 and HepG2 *CYGB*$^{-/-}$ cells were grown at 37 °C in a 5% CO_2 air environment (ESCO, # CCL-170B-8) in DMEM high glucose medium (PAN-Biotech, # P04–03590) complemented with 10% fetal bovine serum (FBS GOLD, # A15–151; GE Healthcare/PAA) and 1% penicillin/streptomycin (PAN, # P06–07100). The editing of HepG2 *CYGB*$^{-/-}$ cell lines is detailed in the Supplementary methods section.

2.8. RNA isolation and cDNA synthesis

Total RNA extraction was performed on 80–90% confluent HepG2 cells from a 9.6 cm^2 plate or from a 25 cm^2 flask using the RNeasy® mini Kit (QIAGEN, # 74104), following the manufacturer's protocol, and including a DNaseI (QIAGEN, # 79254) digestion step. The quality and integrity of the extracted RNA was assessed using an Agilent 2100 Bioanalyzer and the Agilent RNA 6000 Nano Kit (Agilent, # 5067–15). RNA integrity number (RIN) values ranged from 9 to 9.8 indicating high quality and integrity of the RNA samples. RNA concentrations were quantified by fluorometry using the Qubit RNA HS Assay Kit (Invitrogen, # Q32852). cDNA samples for reverse-transcriptase polymerase chain reaction (RT-PCR) and quantitative RT-PCR (qRT-PCR) analyses were synthesized starting from 1 μg of total RNA spiked with 100 ng of total adult *Drosophila melanogaster* RNA using the SuperScript® III Reverse Transcriptase (Invitrogen, # 18080044) following the manufacturer's protocol.

2.9. RT-PCR analyses

RT-PCR experiments designed to investigate the inclusion of the putative exons U_1 and U_2 in *CYGB* transcript variants were performed on RNA samples from HepG2 cells using the U_1, U_2, or *CYGB* V-3 qRT-PCR (cDNA synthesis positive control) forward primers in combination with the *CYGB* qRT-PCR reverse primer. The three reactions were carried out using 1.25 U of Taq DNA Polymerase enzyme (Thermo scientific™, # EP0613), a cDNA volume corresponding to 50 ng of the starting total RNA, and 200 nM of each respective primer in a final volume of 50 μl. The PCR amplification reactions were performed as it follows: an initial denaturation step (5 min at 95 °C) was followed by 40 cycles of denaturation (1 min at 95 °C), annealing (1 min at 56 °C), and elongation (90 s at 72 °C), and completed by a final elongation step (10 min at 72 °C). To validate the *CYGB* knockout at the mRNA level in the selected HepG2 lines, a specific PCR assay was designed to amplify the deleted region (*CYGB* deletion mRNA forward and reverse primers). As positive control, an assay amplifying the unedited 5′ region of the *CYGB* V-3 transcript (*CYGB* V-3 qRT-PCR forward primer and *CYGB* qRT-PCR reverse primer) was used to assess RT-PCR efficiency. PCR reactions were performed as described above with minor variations: amplification was carried out for 35 cycles with both the annealing and elongation steps set to 30 s. 10 μl of PCR products from each sample were resolved by electrophoresis in a 2% agarose gel (Agarose MEEO, Roth # 2268) and visualized by ethidium bromide (Roth, # 7870) staining.

2.10. 5'RACE of CYGB transcripts in HepG2 cells

To investigate the sequence of the 5′ end of the EST-identified *CYGB* V-3 transcript, 5'RACE was performed on RNA from HepG2 cells using the 5'RACE System for Rapid Amplification of cDNA Ends Kit Version 2.0 (Invitrogen, # 18374058), following a slightly modified manufacturer protocol. Briefly, the reverse transcription step was performed on 3 μg of HepG2 RNA (RIN = 9) using SuperScript® III Reverse Transcriptase (Invitrogen, # 18080044) instead of SuperScript® II. The first strand cDNA was synthesized using a *CYGB* specific primer (GSP1) targeting *CYGB* exon B. A first PCR reaction to amplify the 5′-end of the *CYGB* transcripts was performed for 40 cycles of amplification (denaturation step: 1 min at 95 °C; annealing step: 1 min at 62 °C; elongation step: 1 min at 72 °C) on 1 μl of the generated cDNA using 2 U of Taq DNA Polymerase enzyme (Thermo scientific™, # EP0613), 200 nM of 5'Race Abridged Anchor forward primer (Invitrogen, # 10630–010), and 200 nM of a second *CYGB* gene specific reverse primer (GSP2) in a final reaction volume of 50 μl. A nested PCR reaction was performed on 1 μl of products from the first PCR amplification using 1.25 U of Taq DNA Polymerase enzyme (Thermo scientific™, # EP0613) and a nested PCR primer assay consisting of the Abridged Universal Amplification Primer (AUAP, 200 nM) (Invitrogen, # 10541–019) and of a third *CYGB* gene specific reverse primer (GSP3, 200 nM) in a final reaction volume of 50 μl. The nested PCR reaction was performed for 35 cycles (1 min of denaturation at 95 °C, 1 min of annealing at 58 °C, and 1 min of elongation at 72 °C), and 10 μl of the nested PCR products were visualized on

a 1.5% agarose gel. The remaining nested PCR products were purified using the Wizard SV Gel and PCR-Cleanup System (Promega, # A9282) according to the manufacturer's protocol and cloned into 10 ng of pGEM®-T Easy vector (Promega, # A1360) via incubation with 5 U of T4 DNA ligase enzyme (Thermo Scientific™, # EL0011) for 1 h at 22 °C in a final reaction volume of 20 µl. The reaction was heat-inactivated by a 10 min incubation at 65 °C and ligation products were purified by ethanol precipitation, re-suspended in 6 µl of HPLC-H_2O (Roth, # A511.2), and 3 µl were transformed into competent DH10B *E. coli* cells. The transformed bacteria were selected by overnight incubation on LB-ampicillin (100 µg/ml, AppliChem, # A0839) agar plates at 37 °C and positive clones, identified by blue-white screening, were inoculated overnight at 37 °C in 5 ml of LB medium supplemented with 100 µg/ml ampicillin (AppliChem, # A0839) while agitated by a rocking motion at 180 rpm (GFL®, # 3031). Plasmid preparation was performed on 4 ml of liquid bacterial culture using the GeneJet Plasmid preparation Kit (Thermo Scientific™, # K0503) following the manufacturer's protocol. Insert presence was verified by digesting 600 ng of plasmid with 10 U of *Eco*RI restriction enzyme (Thermo Scientific™, # ER0271) for 1 h at 37 °C, followed by gel electrophoresis. Sequencing chromatograms of the *CYGB* 5'RACE amplification products were generated by Sanger sequencing (StarSEQ, Mainz) using SP6 and T7 sequencing primers and they were visualized and analysed using FinchTV (version 1.4.0, Geospiza).

2.11. qRT-PCR analyses

qRT-PCR reactions designed to quantify the transcription of *CYGB* isoforms, *CCNB1*, *CDK1*, *PLK1*, *TOP2A,* and *GAPDH* (as reference gene) in the generated $CYGB^{+/+}$ and $CYGB^{-/-}$ HepG2 lines (named A-C) were performed with GoTaq® qPCR Master Mix (Promega, # A6002) in an Applied Biosystems™ 7500 Fast Real-Time PCR System (Applied Biosystems™) using a cDNA volume corresponding to 17 ng of the starting total RNA and 0.6 µM of the target-specific primer pair in a final volume of 10 µl. Samples were checked for efficiency of reverse transcription via amplification of the spiked-in *Drosophila melanogaster glob1* mRNA. Amplification reactions were performed with standard parameters for 40 cycles using a three-step cycle method (denaturation: 95 °C for 15 s; annealing: 56 °C for 30 s; elongation: 72 °C for 30 s). All qRT-PCR experiments were performed in biological and technical triplicates. Data were analysed using the AB 7500 Software (version 2.0.1, Applied Biosystems™). The absolute copy number/µg of RNA for *CYGB* V-1 and *CYGB* V-3 was calculated using the standard curve method and serial dilutions (10^3–10^7) of known standards for both mRNA isoforms. Relative gene expression was calculated with the $2^{-\Delta\Delta Ct}$ method using the REST software (version 2.0.13) and *GAPDH* as reference.

2.12. Oligonucleotides

All primers and oligonucleotides used in this study were obtained from Sigma-Aldrich as dried and desalted oligonucleotides. Their sequences are listed in Supplementary Table 2.

3. Results

3.1. Bioinformatic identification of putative novel human CYGB mRNA isoforms

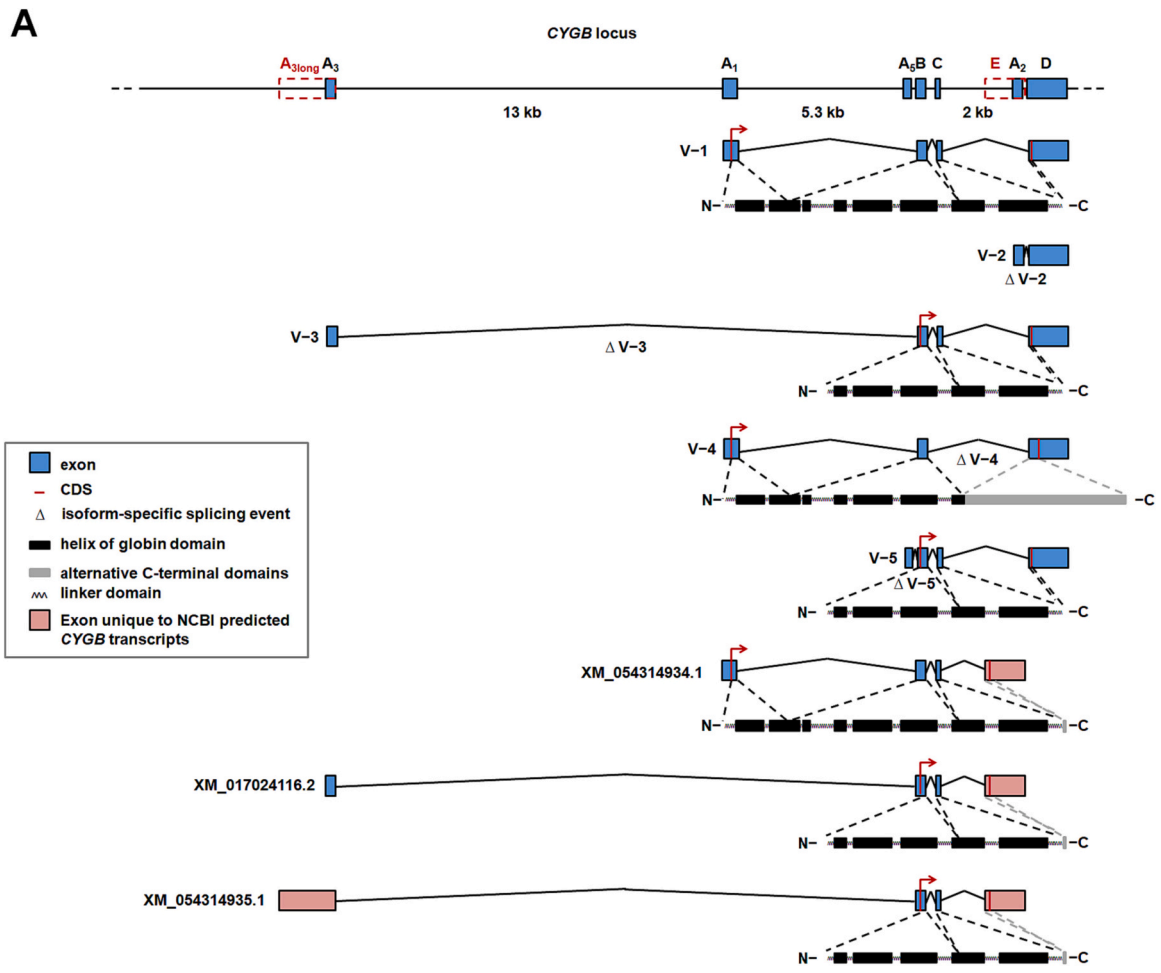
To our knowledge, only one human *CYGB* mRNA isoform has been characterized in literature so far. To investigate whether AS might contribute to the diversity of the proposed CYGB functions, we searched for previously undescribed *CYGB* transcript variants by performing data mining in the human EST (cDNA) collections from the UCSC Genome Browser and AceView databases. EST entries mapping at least partially to any of the known *CYGB* exons were identified and their sequences were analysed for the presence of putative AS events and previously undescribed exonic regions. Four different AS events were found by this approach: three involved distinct alternative first exons and one consisted in an internal exon skipping event. The sequences of the newly identified alternative first exons and of the downstream introns were inspected for the presence of canonical splice sites (GT/AG di-nucleotide sequences [76]) and the AS events could be ascribed to four putative *CYGB* mRNA variants (*CYGB* V-2 to *CYGB* V-5) that are distinct from the canonical *CYGB* V-1 isoform (Fig. 1; Supplementary Fig. 1). Among the analysed ESTs, 28 (originating from normal and tumour brain samples, placenta, eye, spleen, fetal lung, kidney and breast tumour biopsies, BG01 cell line, and embryoid bodies) could be unambiguously attributed to the canonical *CYGB* V-1 isoform. Instead, 5 EST evidences (generated from placenta, a pooled lung and spleen library, lung large cell carcinoma, Jurkat cell line, and a chondrosarcoma lung metastasis cell line) supported the presence of *CYGB* isoform V-3. Contrarily, only one EST evidence was detected for the *CYGB* transcript variants V-4 (from a brain tumour sample), V-2, and V-5 (both from a pooled library of lung and spleen samples), suggesting that these mRNAs might be of less relevance in terms of transcriptional representation (also see 3.3).

Importantly, the five identified *CYGB* mRNAs matched, with some discrepancies at their 3′ and 5′ ends, the *CYGB* isoforms reported in the Ensembl database (http://www.ensembl.org, release 109) (Fig. 1B), which, however, are each supported there only by a single EST evidence. Four *CYGB* transcripts, instead, are included in the NCBI nucleotide database (https://www.ncbi.nlm.nih.gov/nuccore) (Fig. 1): NM_134268.5 corresponds to the RefSeq annotation for the canonical *CYGB* mRNA, XM_054314934.1 is an automatically annotated, predicted transcript matching NM_134268.5, but with a 1.4 kbp-long alternative last exon (E), and XM_017024116.2 and XM_054314935.1 are two predicted transcripts consisting of exons A_3 or A_{3long} (including the ~1.5 kbp of upstream A_3), respectively, B, C, and E. Notably, no cDNA evidence supported these automatically annotated transcripts. Furthermore, in a human HB cell line, CAGE and RNA-PET data (see below) and RT-PCR product sequencing showed expression of *CYGB* V-3, but not of exon E (Supplementary Fig. 2**,** Supplementary Sequence 1). This suggests that the alternative last exon included in the NCBI-predicted *CYGB* variants might represent an artefact, possibly generated by automatic annotation. However, it cannot be excluded that this candidate variant is transcribed in cells or under conditions different from the ones analysed in this study.

3.2. ORF analysis of the CYGB mRNA isoforms and protein structure prediction

To investigate if the identified *CYGB* mRNA isoforms code for canonical or modified CYGB proteins, their sequences were searched for open reading frames (ORFs) (Fig. 1B, Supplementary Fig. 3). *CYGB* V-1 ORF matched the coding sequence (CDS) for the canonical CYGB protein. No ORF was detected in the short *CYGB* V-2 transcript, suggesting that it is not functional. Interestingly, a 378 bp ORF, potentially encoding a truncated CYGB protein lacking the first N-terminal 65 aa of the full-length version, was identified in the sequence of *CYGB* V-3, V-3r (see below), and V-5. The candidate truncated protein retains the conserved His81 and His113 residues, which are crucial for heme and ligand binding in globins, but lacks the Cys38 residue and an evolutionary conserved phenylalanine at position CD1 (Phe60) that is involved in the stabilization of the heme [1,77,78]. Analysis of *CYGB* V-4 identified an ORF potentially coding for a 226 aa protein whose 125 N-terminal residues match those of canonical CYGB. This candidate protein retains the biologically meaningful residues His81, His113, Cys38, Cys83, and Phe60, but has a C-terminus that differs from canonical CYGB. To investigate whether the predicted ORFs could be translated, the sequences encompassing the start codons were searched for the presence of Kozak sequences (5'-GCCRCCaugG-3′) [79–81]. Non-canonical Kozak sequences were identified for all predicted ORFs (5'-CUGCUCaugG-3′ for *CYGB* V-1 and *CYGB* V-4, 5'-AAGCACaugG-3′ for

A



B

Transcript name / ID	Database	Length (bp)	Exons	Protein	Expression
CYGB V-1	this study	2,158	A_1(506bp); B(232bp); C(164bp); D(1256bp)	full-length CYGB (190aa)	broad array of tissues and cells
CYGB V-2	this study	1,540	A_2(284bp); D(1256bp)	not protein coding	lung; spleen
CYGB V-3	this study	1,966	A_3(314bp); B(232bp); C(164bp); D(1256bp)	N-terminally truncated CYGB (125aa)	liver; lung; lymph node; placenta; spleen; mammary epithelial cells; hepatoblastoma tissue and cell lines; lung large cell carcinoma; chondrosarcoma lung metastasis cell line; Jurkat cell line; SNU-719 cell line
CYGB V-4	this study	1,770	A_1(282bp); B(232bp); D(1256bp)	V-4 CYGB isoform (226aa)	brain tumour
CYGB V-5	this study	1,843	A_5(191bp); B(232bp); C(164bp); D(1256bp)	N-terminally truncated CYGB (125aa)	lung; spleen; keratinocytes
ENST00000293230.10	Ensembl	1,962	A_1(310bp); B(232bp); C(164bp); D(1256bp)	full-length CYGB (190aa)	broad array of tissues and cells
ENST00000586160.1	Ensembl	842	A_2(287bp); D(555bp)	not protein coding	lung; spleen
ENST00000589145.1	Ensembl	788	A_3(314bp); B(232bp); C(164bp); D(78bp)	N-terminally truncated CYGB (125aa)	brain; dorsal root ganglia; lung; spleen; kidney tumor; neuroblastoma; Jurkat cell line
ENST00000589342.1	Ensembl	815	A_1(253bp); B(232bp); D(330bp)	V-4 CYGB isoform (226aa)	brain; dorsal root ganglia; kidney tumor; neuroblastoma
ENST00000590175.5	Ensembl	830	A_5(190bp); B(232bp); C(164bp); D(244bp)	N-terminally truncated CYGB (125aa)	brain; dorsal root ganglia; lung; spleen; kidney tumor; neuroblastoma
NM_134268.5	NCBI	1,962	A_1(310bp); B(232bp); C(164bp); D(1256bp)	full-length CYGB (190aa)	broad array of tissues and cells
XM_054314934.1	NCBI	2,108	A_1(310bp); B(232bp); C(164bp); E(1402bp)	XP_005257062 (202aa)	predicted transcript automatically annotated by computational analysis of genomic sequence
XM_017024116.2	NCBI	2,132	A_3(332bp); B(232bp); C(164bp); E(1404bp)	XP_016879605 (137aa)	predicted transcript automatically annotated by computational analysis of genomic sequence
XM_054314935.1	NCBI	3,585	A_{3long}(1787bp); B(232bp); C(164bp); E(1402bp)	XP_016879605 (137aa)	predicted transcript automatically annotated by computational analysis of genomic sequence

Fig. 1. Complexity of the human *CYGB* gene and its transcripts. A) Schematic representation of the human *CYGB* gene locus, its predicted mRNA isoforms as inferred by cDNA data mining (V-1 to V-5), and the predicted *CYGB* transcripts reported in the NCBI database. The simplified secondary structures of the canonical CYGB protein (V-1) and the putative alternative CYGB isoforms (V-3/V-5, V-4, XP_005257062, and XP_016879605) are reported underneath the respective transcripts. Block boxes depict the globin alpha helices, vertical red arrows and red lines indicate start and stop codons, respectively. A, B, C, D (blue squares) are exons of the identified *CYGB* transcripts with A_1, A_2, A_3, A_5, being the first exon of isoforms V-1and V-4, V-2, V-3, and V-5, respectively. A_{3long} and E (red boxes) represent exons unique to the predicted *CYGB* transcripts included in NCBI database. B) Table summarizing and comparing the *CYGB* transcripts described in our study and the ones reported in the Ensembl and NCBI nucleotide databases, including the reported expression sites of the transcripts and their supporting evidences. Note that the most robust expression data are present for V-1 and V-3. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

CYGB V-3 and *CYGB* V-5).

The predicted 3D structures of all CYGB isoforms are documented in the AlphaFold protein structure database (Fig. 2A-C). Pairwise structure analysis showed that the first 116 aa of the truncated V-3 isoform had an almost identical predicted structure to its respective region of full-length V-1 CYGB (Fig. 2D). The V-4 structure, however, diverged from both other isoforms (Fig. 2E and F), since it contains large unstructured parts, predicted with a very low confidence score (Fig. 2C). No predicted 3D structure, instead, is available for the NCBI-predicted CYGB isoforms XP_005257062 (202 aa) and XP_016879605 (137 aa), which are potentially encoded by XM_054314934.1 and XM_017024116.2 and XM_054314935.1, respectively (Fig. 1B). These predicted proteins would match the first 180 aa and 115 aa of the full-length V-1 and truncated V-3 CYGB, respectively, and have an identical 22 aa-long C-terminus.

3.3. Expression analysis of the CYGB isoforms in human tissues and cells

In a large transcriptomics-based re-evaluation of *CYGB* expression in mammalian organisms, over 200 publicly available RNA-seq datasets from human normal tissues, primary cells, and cancer cell lines were analysed and the expression of the *CYGB* gene for each sample was measured by RPKM values [72]. Transcriptome analyses were performed using a human reference genome version that included annotations for all *CYGB* mRNA isoforms. For every RNA-seq dataset that showed transcription of *CYGB* (RPKM>0.5), the percentage of each *CYGB* isoform was estimated by calculating the proportion of reads mapping to the isoform-specific splicing events (see Supplementary methods). As expected, *CYGB* V-1 was the most abundant *CYGB* mRNA isoform across tissues, primary cells, and cell lines (>90% in the majority of the samples) (Supplementary Tables 3 and 4). V-3 was observed to be the dominant *CYGB* mRNA isoform in the gastric cancer cell line SNU-719 (99–100%) and in multiple human HB cell lines: HepG2 (94–100% $n = 7$, confirmed by qRT-PCR, Supplementary Fig. 4), Hep293TT (78–100% $n = 2$), HepU2 (100% n = 2), and Huh-6 (100% n = 2). Remarkably, *CYGB* isoform V-3 was also detected in biopsies from normal liver (0–40% n = 2), lung (9–31% n = 2), lymph node (25% $n = 1$), and in primary mammary epithelial cells (HMEC, 34–58% n = 2). Limited transcription (<15%) was observed for the other *CYGB* isoforms with the exception of *CYGB* V-5 in primary normal epidermal keratinocytes (NHEK, 19–25%, $n = 3$).

In light of the detection of *CYGB* V-3 mRNA in normal liver biopsies

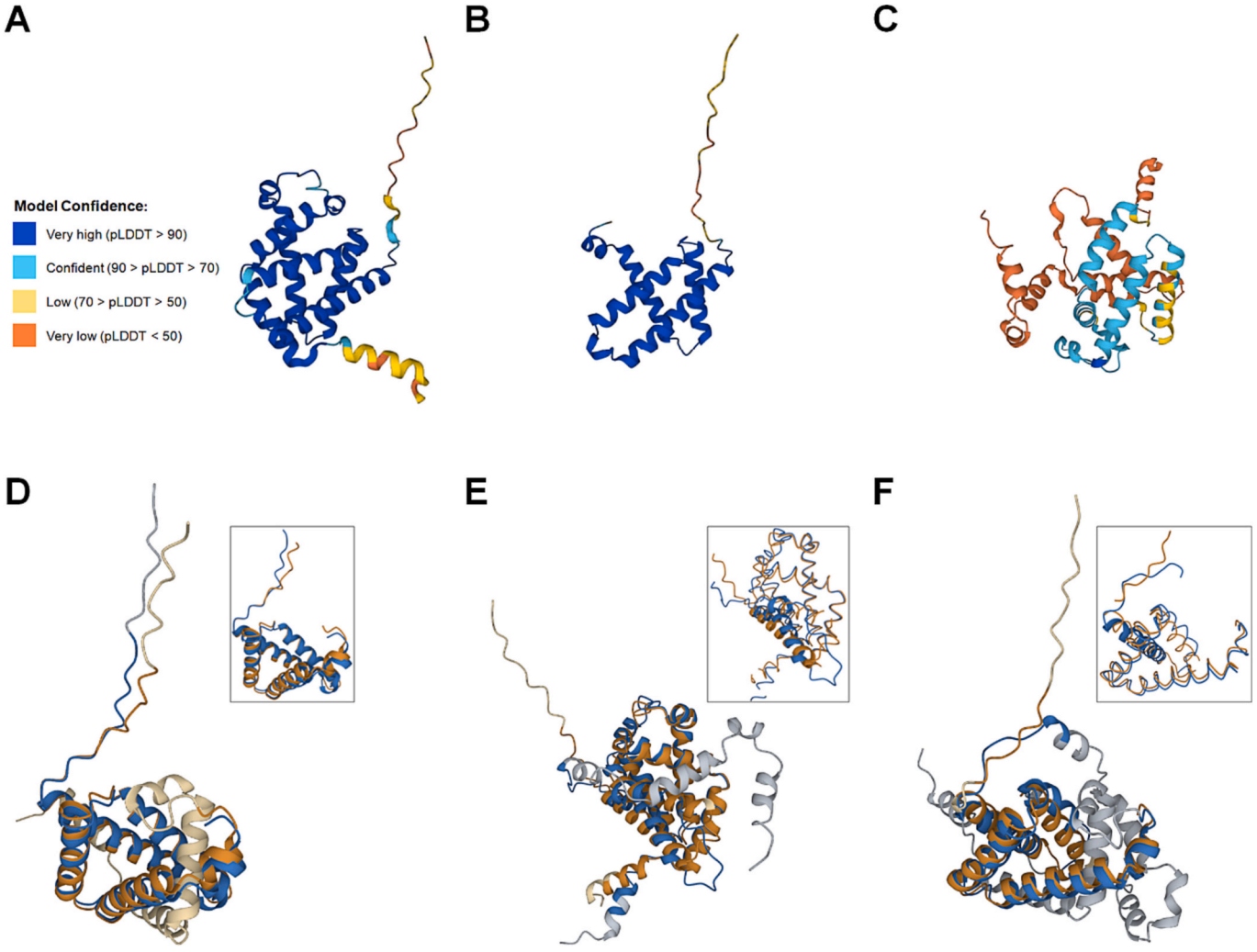


Fig. 2. CYGB isoforms: structure prediction and comparison. A-B) 3D structure predictions for the canonical V-1 (Q8WWM9) (A), truncated V-3 (K7EIM9) (B), and V-4 (K7EMC7) (C) CYGB protein isoforms were obtained from AlphaFold database (https://alphafold.com/) [65,66]. D-F) Pairwise structure alignments of the canonical V-1 CYGB (orange) and the truncated V-3 (D) or V-4 (E) isoforms (blue) and of the truncated V-3 (orange) and V-4 (blue) isoforms (F). The aligned residues for each comparison are reported in boxes. The truncated V-3 protein folds almost identically to its respective region in full-length V-1 CYGB, while the predicted structure of V-4 differs from the other CYGB isoforms. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

and HB cell lines, we analysed publicly available RNA-seq datasets from a study including normal and HB liver samples of the subtypes C1 (T-C1) and C2 (T-C2). *CYGB* expression was detected in all samples from normal and HB liver tissues. Differential gene expression analysis showed that T-C1 tissues had higher overall *CYGB* mRNA levels compared to normal tissues and T-C2 samples. No significant difference in the overall *CYGB* expression was observed between normal tissues and T-C2 tissues (Fig. 3, Supplementary File 2). At the isoform level, V-1 and V-3 were estimated to be the predominant *CYGB* transcript variants in all analysed samples. V-1 was found to be the major isoform in 16.7% of normal liver tissues (4/24), in 16.7% of T-C1 samples (3/18), and in 83.3% of T-C2 samples (5/6). Conversely, *CYGB* V-3 was estimated to be the principal isoform in 83.3% of normal liver tissues (20/24), in 83.3% of T-C1 samples (15/18), and in 16.7% of T-C2 samples (1/6). V-1 accounted for 7–100% of the *CYGB* mRNAs across normal liver tissues, as well as for 5–77% and 35–97% of the *CYGB* transcripts in T-C1 and T-C2, respectively. Instead, *CYGB* V-3 represented 0–89%, 21–95%, and 0–65% of the *CYGB* transcripts in normal, T-C1, and T-C2 samples, respectively (Supplementary Table 5).

Altogether, our transcriptome analyses confirmed that *CYGB* V-1 is the major *CYGB* mRNA isoform in human tissues and cells. Furthermore, the remarkable abundance of *CYGB* V-3 in both normal and HB liver samples suggests a functional role for this *CYGB* transcript in these tissues.

3.4. Promoter prediction and regulatory sequence analyses

The identification of alternative first exons in *CYGB* transcripts advocates for the presence of alternative promoters. Therefore, we searched the genomic sequence of the *CYGB* locus plus 20 kb of its upstream and downstream regions for possible promoter regions using the Gene2Promoter tool of Genomatix. The analysis identified three putative promoter regions: GXP_55881, GXP_4415812, and GXP_4415813 (Supplementary Fig. 1B). GXP_55881 coincides with the previously described promoter of the canonical transcript isoform *CYGB* V-1 [48,53]; furthermore, because *CYGB* V-1 and V-4 share the same first exon (A_1), GXP_55881 could potentially also regulate the transcription of *CYGB* V-4. A candidate promoter GXP_4415812 was reported upstream of *CYGB* exon D, partially overlapping the first alternative exon A_2. Another candidate, GXP_4415813, spans exon A_5 and the most 5′ region of *CYGB* exon B, which includes the predicted transcription starting site (TSS) of isoform V-5. In contrast, the analysis failed to identify a potential promoter that could account for the transcription of *CYGB* V-3 in the genomic sequence encompassing exon A_3 or in its proximity. However, a putative promoter region, GXP_4415818, was identified ~6.4 kbp upstream of exon A_3 on the correct (−) strand. This candidate promoter partially overlaps two CpG islands (692 bp and 221 bp in length, respectively; Supplementary Fig. 1C), which are common features of mammalian promoters [82]. Via EST database mining, we identified 4 spliced EST evidences mapping on the (−) strand of the GXP_4415818 locus (Supplementary Fig. 1F). Importantly, three of these ESTs (DB034636, HY007991, and BF726740) mapped also to *CYGB* exon A_3, while the fourth one (DB244936, originating from uterus cancer tissues) mapped to a 510 bp region located ~400 bp upstream of *CYGB* exon A_3 (Supplementary Fig. 1F). Two of the EST evidences (DB034636 and HY007991), both generated from a testis cDNA library, showed an identical splicing pattern characterized by three exonic regions: the first exon (U_1) is located in the GXP_4415818 promoter region, the second is *CYGB* exon A_3, and the third is found ~5 kbp downstream of *CYGB* exon A_3. Finally, one EST evidence (BF726740), which was generated from a library of human lens mRNA, displayed a different 3-exon pattern consisting of U_1, a second exon located ~300 nt downstream of it, and *CYGB* exon A_3. The identification of EST evidences mapping to both GXP_4415818 and *CYGB* exon A_3 indicated the presence of additional *CYGB* upstream exons whose transcription might be regulated by the predicted promoter region GXP_4415818. By confirmatory RT-PCR and 5'RACE experiments we identified two populations of *CYGB* mRNA molecules in HepG2 cells: one corresponding to isoform V-3, and the other to a *CYGB* transcript including *CYGB* exon A_3 and the upstream exon U_1 (Supplementary results and Supplementary Figs. 1E, 5–7). For simplicity, we referred to this novel mRNA isoform as *CYGB* V-3r and to exon U_1 as exon A_{3R}.

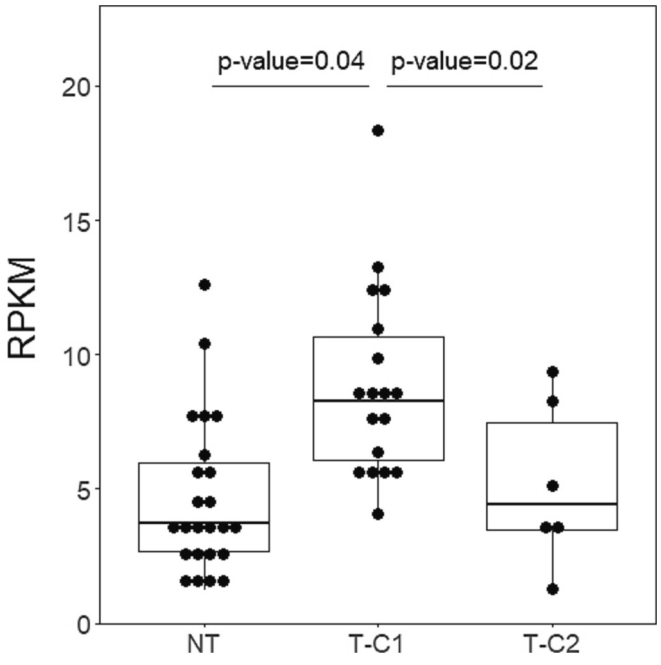


Fig. 3. *CYGB* transcription in tissues from normal livers and HB subtypes. RNA-seq and differential gene expression analyses indicate that samples of HB subtype C1 have significantly higher overall *CYGB* mRNA levels compared to the normal liver tissues. NT: non-tumourous liver; T-C1: HB subtype C1; T-C2: HB subtype C2; *p*-value: FDR-adjusted p-value.

The sequences of the identified putative promoters were further searched for the presence of TATA, GC, and CAATT boxes. Confirming literature [48], no TATA box was identified in promoter region GXP_55881. Similarly, no TATA box was found in the sequences of the putative promoters GXP_4415818, GXP_4415813, and GXP_4415812 (Supplementary Fig. 8). In line with previous reports [48], multiple GC boxes were observed in promoter region GXP_55881; moreover, 4 potential GC boxes were detected in the sequence of the predicted promoter GXP_4415818, 2 in GXP_4415812, and 1 in GXP_4415813, while no CAATT box was found in any of the analysed sequences (Supplementary Fig. 8).

3.5. Regulation of CYGB expression in HepG2 cells

Our RNA-seq and qRT-PCR analyses showed that V-3 is the predominant *CYGB* isoform in the HepG2 cell line. To investigate the regulation of the V-3 isoform, the UCSC Genome Browser was used to search for markers of open chromatin state and active transcription in the *CYGB* gene locus of HepG2 cells (Fig. 4). In line with our expression data, which detected low-to-no transcription of *CYGB* V-1 in these cells, DNA methylation tracks showed that the CpG island encompassing the GXP_55881 promoter region was hypermethylated in HepG2 cells (Fig. 4, red vertical bars). Instead, no methylation was found in the CpG islands encompassing the putative promoter region GXP_4415818 (Fig. 4, green vertical bars). Furthermore, DNaseI hypersensitivity and FAIRE-seq tracks identified open chromatin segments in the genomic regions of the putative promoter GXP_4415818 and of *CYGB* exon A_3, but not of the GXP_55881 promoter. Importantly, data from multiple RNA polymerase II (Pol II) ChIP-seq experiments indicated the presence of high peaks in the GXP_4415818 region (Fig. 4, dark gray boxes), of

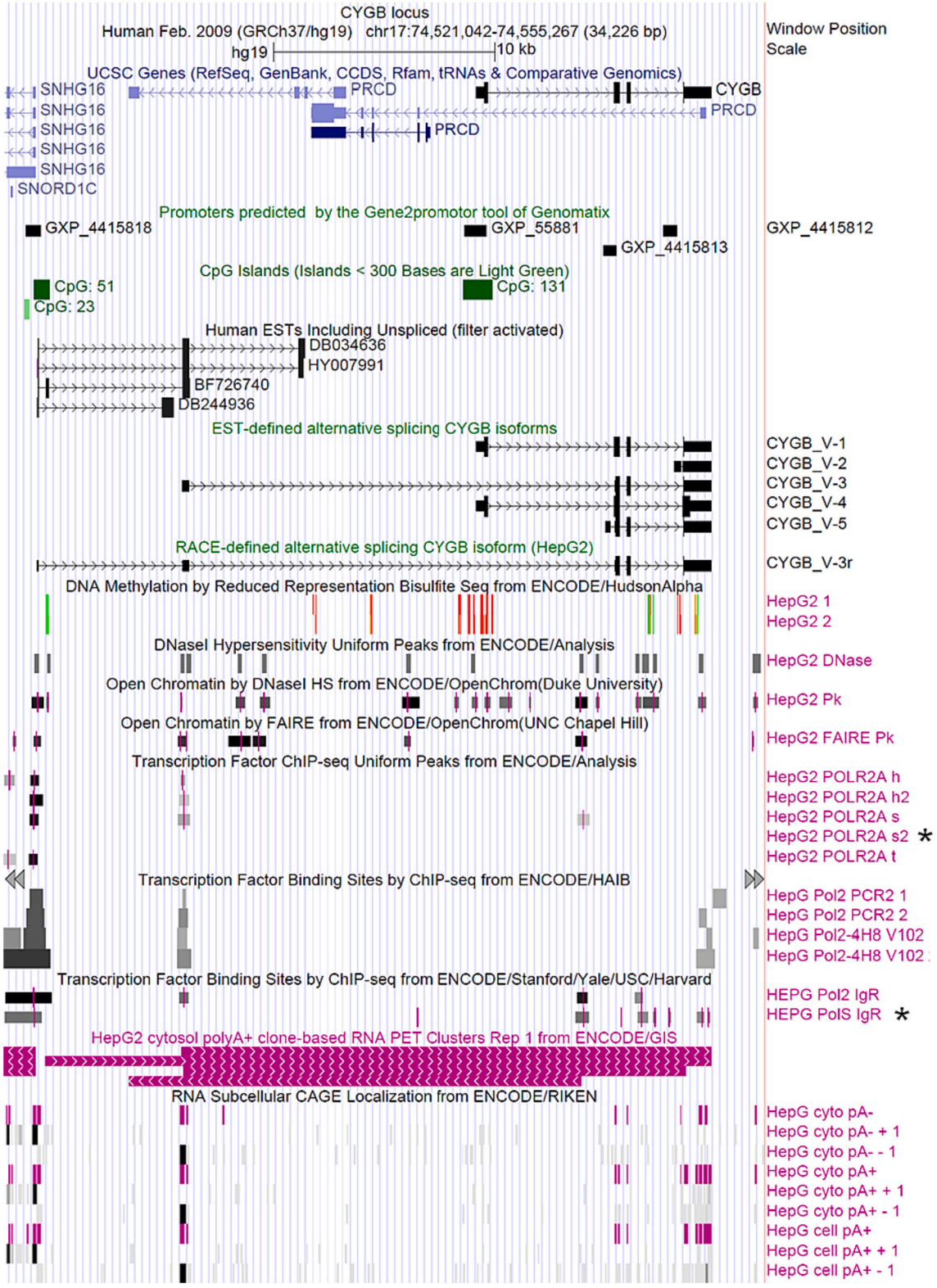


Fig. 4. Markers of active *CYGB* transcription in the HepG2 cell line. UCSC Genome Browser visualization of *CYGB* gene structure, *CYGB* mRNA expression and transcriptionally active sites in HepG2 cells. Most publicly available data are from ENCODE (www.encodeproject.org). Custom data tracks added by us are indicated below by a hashtag (#). From the top, the following tracks are reported: UCSC Genes; putative promoter regions predicted by the Gene2Promoter tool[#]; CpG islands; EST evidences mapping to the putative GXP_4415818 promoter region[#]; EST-defined *CYGB* mRNA isoforms[#]; 5' RACE-defined *CYGB* V-3r isoform[#]; DNA methylation tracks (green indicates no methylation, red indicates hypermethylation); DNaseI hypersensitivity tracks; FAIRE-seq tracks; Pol II ChIP-seq tracks (the grayscale shade is proportional to the height of the peaks); RNA-PET tracks; CAGE tracks. Data are represented on the Crick strand. For each CAGE track the total mapping reads as well as the reads mapping only to the Watson strand (+1) or only to the Crick strand (−1) are reported (the gray-scale shade is proportional to the number of mapping reads). *: Pol II phosphorylated Ser2 ChIP-seq tracks. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

lower peaks in the genomic region of *CYGB* exon A_3 (Fig. 4, light gray boxes), and the absence of peaks in any of the other validated or putative *CYGB* promoter regions. However, data from the two reported Ser2 phosphorylated Pol II ChIP-seq experiments, which are designed to identify sites of active transcription initiation, indicated the absence of peaks in the genomic region of *CYGB* exon A_3, while tracks from only one of these experiments showed a peak in the GXP_4415818 region. In line with the DNA-based data, evidences from cap analysis of gene expression (CAGE) and paired-end diTag sequencing of RNA (RNA-PET) experiments confirmed the expression of *CYGB* V-3, but not of *CYGB* V-1 in HepG2 cells; thus, these data indicated the synthesis of a putative *CYGB* mRNA isoform whose transcription starts immediately downstream of the putative GXP_4415818 promoter. Altogether, RNA-based data (ESTs, CAGE, and RNA-PET) and DNA-based data (DNA methylation, DNaseI hypersensitivity, FAIRE-seq, and Pol II ChIP-seq) suggested that in HepG2 cells *CYGB* V-1 is not expressed, while the transcription of putative alternative *CYGB* mRNA isoforms might be initiated both from the predicted promoter region GXP_4415818 and from the A_3 exonic region.

3.6. Conservation of CYGB transcript isoforms

The conservation of an alternative mRNA isoform across different organisms is generally considered as an indication of its biological relevance [83]; therefore, we investigated whether sequences homologous to the EST- and 5'RACE-identified *CYGB* isoforms are found in organisms other than humans via sequence searches. EST mining failed to identify evidence supporting the expression of transcripts homologous to the *CYGB* isoforms, except for the canonical V-1 variant (data not shown). Because the available EST data were derived from a limited number of organisms and tissues, we further searched the NCBI nucleotide database, which includes manually curated and automatically annotated entries, for transcripts homologous to the alternative human *CYGB* isoforms (Table 1). Mining of this database failed to identify sequences homologous to *CYGB* V-2; however, predicted *Cygb* transcript variants from *Cebus capucinus imitator* (X1, X2, X3, and X5) are reported to include an exonic region homologous to a shortened *CYGB* exon A_2. Similarly, no homologous transcript was identified for *CYGB* V-5, but multiple predicted *Cygb* transcript variants from *Aotus nancymaae* (X3, X4, X5, and X6) and *Cebus capucinus imitator* (X5) included exon regions homologous to *CYGB* exon A_5. Transcripts homologous to *CYGB* V-3 were identified among the predicted mRNAs of *Cercocebus atys* (X6), *Mandrillus leucophaeus* (X3), *Papio anubis* (X5), and *Theropithecus gelada* (X2). Finally, transcripts homologous to *CYGB* V-4 were detected in *Microtus ochrogaster* (X2) and *Pan troglodytes* (X2).

3.7. Establishment of $CYGB^{-/-}$ HepG2 cell lines

To gain insights into the possible functions of the *CYGB* mRNA isoforms V-3 and V-3r, we generated *CYGB* knockout ($CYGB^{-/-}$) HepG2 cell lines using CRISPR/Cas9n technology with a modified humanized Cas9 nickase enzyme to minimize potential off-target effects. The Cas9n gene and two gRNAs designed to target *CYGB* exon B were transiently transfected into HepG2 cells alone or together with a pair of gRNAs targeting the 3'end of *CYGB* exon C. When transfected alone, the Cas9n gene and the first gRNA pair generated short deletions that encompass the functionally important His81 residue; instead, the combination of the two gRNA pairs created long deletions that resulted in the excision of His81 and the likewise crucial residue His113. Importantly, both the short and the long deletions should preclude the synthesis of any functional CYGB protein in the edited $CYGB^{-/-}$ HepG2 cell lines (Fig. 5). The *CYGB* knockout was validated in the three $CYGB^{-/-}$ lines at the DNA and mRNA levels by genotyping PCR and Sanger sequencing (Fig. 5, Supplementary Fig. 9) and by RT- and qRT-PCR analyses, respectively (Fig. 6).

Western blot (WB) analyses were performed using a published and often-used anti-Cygb antibody, which recognizes the first N-terminal 18 aa of the canonical rat Cygb protein [2] and thus should not detect the truncated V-3 hypothetical protein. This antibody detected a strong band at 21 kDa in protein extracts from cell lines expressing the canonical *CYGB* V-1 transcript, but unexpectedly also highlighted the same band in cell types not expressing *CYGB* mRNA and, moreover, in the V-3-expressing $CYGB^{+/+}$ and $CYGB^{-/-}$ HepG2 cells (Supplementary results and Supplementary Fig. 10). Possibly, this anti-Cygb antibody did not specifically recognize the human CYGB protein in our setting, but rather coincidentally detected another protein of same size.

Table 1
Conservation of *CYGB* transcript isoforms and exon regions. The NCBI nucleotide database was searched for transcripts homologous to *CYGB* isoforms. Sequences homologous to the full-length mRNA alternative isoforms (V-3 and V-4), or to *CYGB* alternative first exon regions were identified mainly in transcripts from primate organisms.

Organism	Transcript ID	Transcript name	Sequence homologoy to human *CYGB* transcript isoform or exonic region
Aotus nancymaae	XM_012453647.2	predicted *Cygb* transcript variant X3	A_1, A_5, B, C
Aotus nancymaae	XM_012453648.2	predicted *Cygb* transcript variant X4	A_5, B, C
Aotus nancymaae	XM_012453649.2	predicted *Cygb* transcript variant X5	A_1, A_5, B, C
Aotus nancymaae	XM_012453650.2	predicted *Cygb* transcript variant X6	A_1, A_5, B, C
Cebus capucinus imitator	XR_005201434.1	predicted *Cygb* transcript variant X1	A_1, B, C, A_2 (short), D
Cebus capucinus imitator	XR_005201435.1	predicted *Cygb* transcript variant X2	A_1, B, C, A_2 (short), D
Cebus capucinus imitator	XR_005201436.1	predicted *Cygb* transcript variant X3	A_1, B, C, A_2 (short), D
Cebus capucinus imitator	XR_005201437.1	predicted *Cygb* transcript variant X5	A_1, A_5, B, C, A_2 (short), D
Cercocebus atys	XM_012042359.1	predicted *Cygb* transcript variant X6	*CYGB* V-3
Mandrillus leucophaeus	XM_011984757.1	predicted *Cygb* transcript variant X3	*CYGB* V-3
Microtus ochrogaster	XM_026780441.1	predicted *Cygb* transcript variant X2	*CYGB* V-4
Pan troglodytes	XM_016932957.3	predicted *Cygb* transcript variant X2	*CYGB* V-4
Papio anubis	XM_017950996.3	predicted *Cygb* transcript variant X2	A_3, B, C
Papio anubis	XM_031657799.1	predicted *Cygb* transcript variant X3	A_3, B, C
Papio anubis	XM_017950997.3	predicted *Cygb* transcript variant X5	*CYGB* V-3
Theropithecus gelada	XM_025363619.1	predicted *Cygb* transcript variant X2	*CYGB* V-3

3.8. Phenotypic effects of CYGB depletion in HepG2 cells

The identification of multiple *CYGB* transcript isoforms prompted us

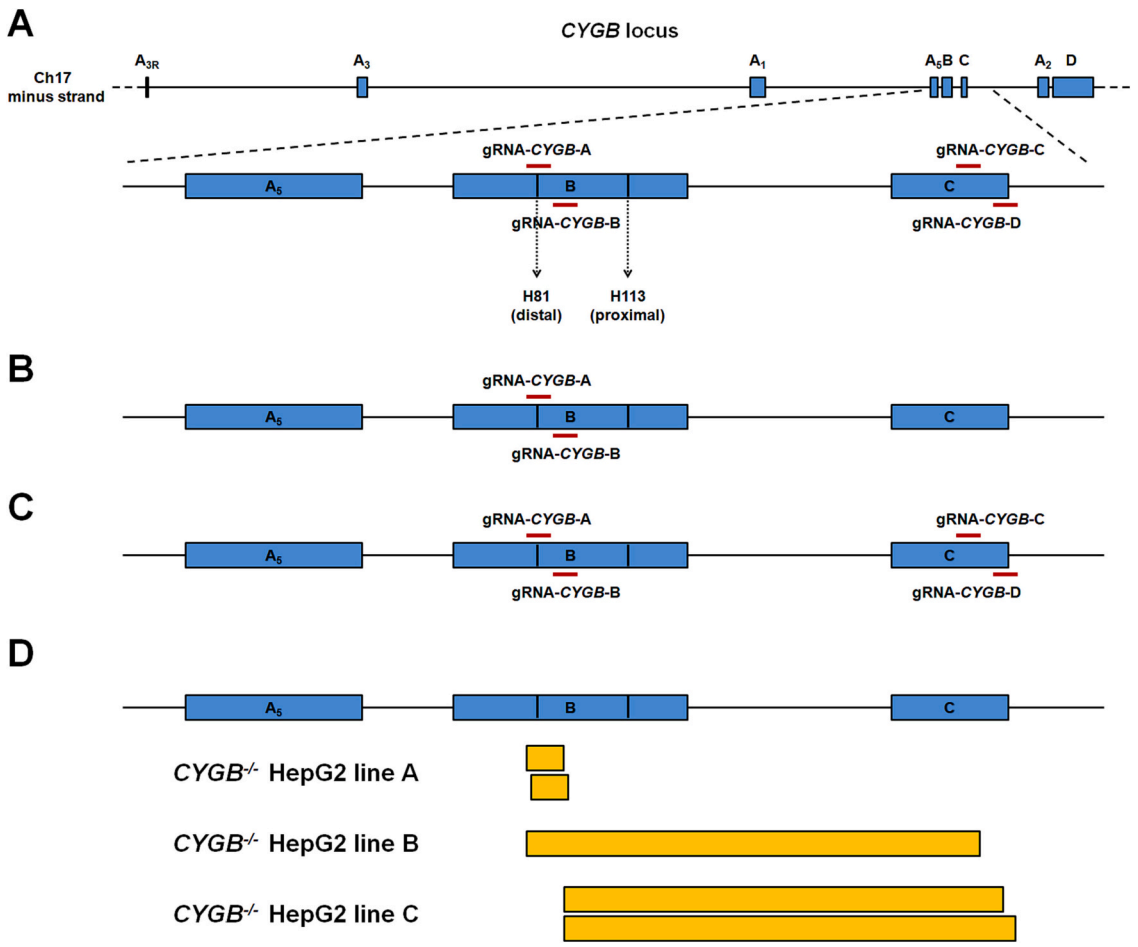


Fig. 5. Generation of *CYGB*$^{-/-}$ HepG2 cell lines via CRISPR/Cas9n technology. A) Schematic representation of the human *CYGB* gene locus and of the *CYGB*-specific gRNAs (red horizontal lines). Vertical black lines indicate the functionally important histidine codons H81 and H113. B-C) Experimental design used to generate short (B) or long (C) deletions in the *CYGB* gene locus. D) The genomic regions deleted in the generated *CYGB*$^{-/-}$ HepG2 cell lines, named A-C, are graphically represented as orange boxes. In *CYGB*$^{-/-}$ HepG2 lines A and C different allelic deletions were observed. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

to evaluate their proportion in transcriptome data from a human kidney podocyte cell model which focused on CYGB's role in chronic kidney disease [39]. In this study, *CYGB* transcription in the human podocyte cell line AB8/13 was triggered by induction of cell differentiation [39]. Furthermore, a *CYGB* knockdown in differentiated AB8/13 cells induced deregulation of genes involved in oxidative stress response, apoptosis, and cell cycle regulation. Among those, regulators of the G2/M checkpoint cyclin B1 (*CCNB1*) [84], cyclin dependent kinase 1 (*CDK1)* [85], polo-like kinase 1 (*PLK1*) [86], and DNA topoisomerase IIα (*TOP2A*) [87] were significantly downregulated in *CYGB*$^{-/-}$ AB8/13 compared to *CYGB*$^{+/+}$ AB8/13 cells. *Re*-investigation of the respective transcriptome data indicated that, even if isoform V-1 was expressed, V-3 was in fact the major *CYGB* transcript variant in differentiated AB8/13 cells.

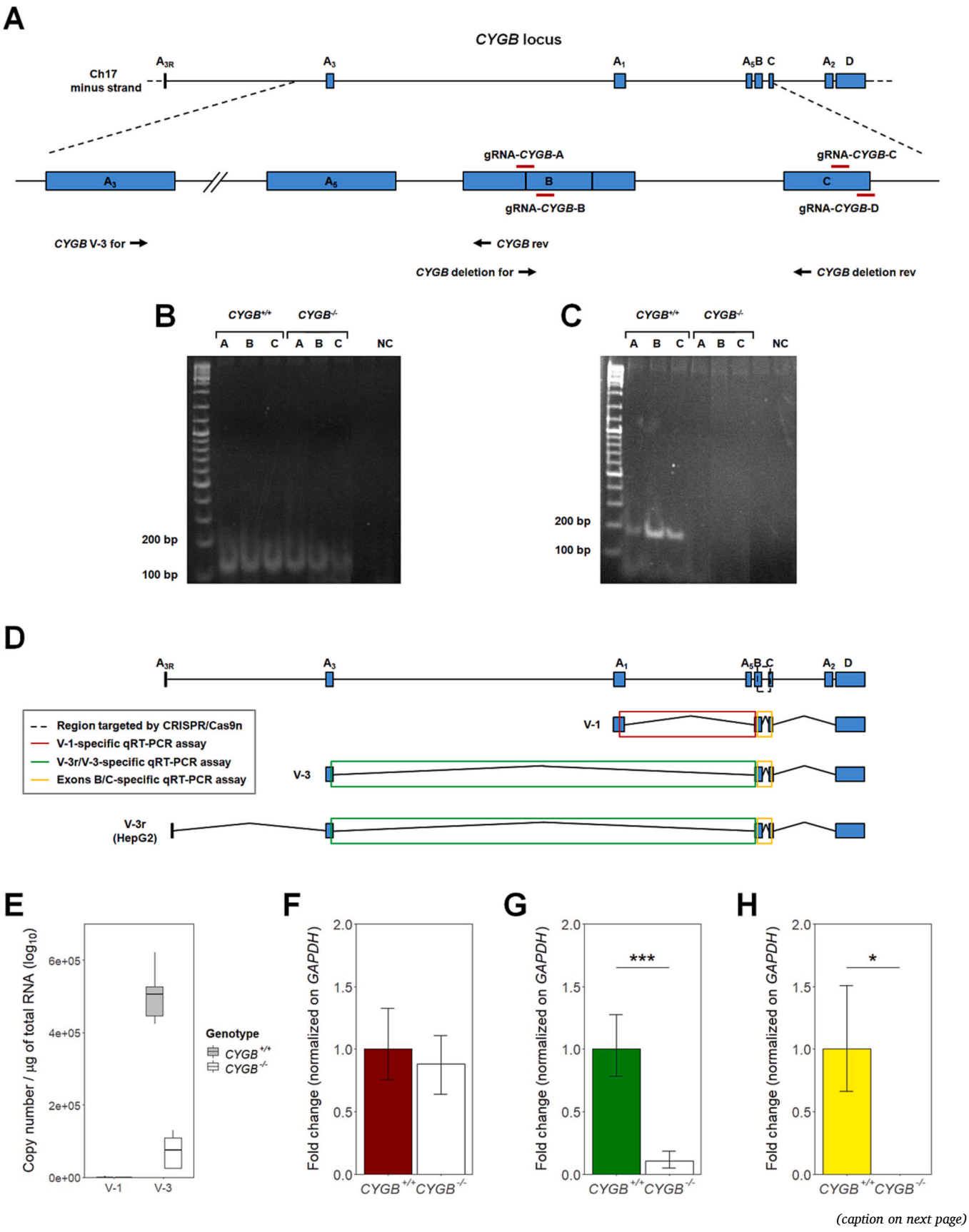
In light of these data, we investigated whether *CYGB* V-3 was involved in cell cycle regulation in our HepG2 model. However, qRT-PCR analyses showed that the *CYGB* knockout in HepG2 cells only resulted in small transcriptional downregulation of *TOP2A*, but did not induce significant changes in the expression of *CCNB1*, *CDK1*, or *PLK1* (Fig. 7). In line with this, flow cytometry did not show differences in the cell cycle profile of *CYGB*$^{+/+}$ and *CYGB*$^{-/-}$ HepG2 cells (Supplementary Fig. 11). To systematically investigate possible transcriptomic effects of the *CYGB* knockout, we performed RNA-seq on *CYGB*$^{+/+}$ and *CYGB*$^{-/-}$ HepG2 cells followed by differential gene expression analysis (see Supplementary methods; ENA accession PRJEB70417). Surprisingly, the transcriptome analysis showed no prominent molecular phenotype, as no gene was significantly dysregulated. This might argue against a functional role of *CYGB* V-3 in HepG2 cells, at least under the applied conditions (Fig. 8, Supplementary File 3). Of note, differential gene expression analysis showed a limited and not statistically significant downregulation for the *CYGB* gene itself. This was explained by the transcription of the remnant *CYGB* exonic regions (Fig. 8D), as also observed via RT-PCR and qRT-PCR analysis (Fig. 6).

4. Discussion

In vitro and *in vivo* studies have proposed multiple physiological roles for Cygb (see Introduction), suggesting that this globin might exert different functions depending on its expression site, its subcellular localization, and the cellular redox state. Since AS, the process by which a single gene produces functionally different mRNAs and proteins [88,89], is widespread in human genes [62], we aimed at investigating whether AS could contribute to the diversity of functions proposed for Cygb.

4.1. Alternative CYGB transcript isoforms: Architecture of the CYGB gene locus and evolutionary conservation

By transcriptome data mining, we identified four putative alternative mRNA isoforms for the human *CYGB* gene (*CYGB* V-2 to V-5) that are distinct from the well-characterized canonical *CYGB* V-1 transcript. The



(caption on next page)

Fig. 6. Validation of the *CYGB* knockout in HepG2 cell lines at the mRNA level. A) Schematic representation of the *CYGB* genomic locus and of the RT-PCR assays used to investigate the presence of *CYGB*-specific deletions in the generated $CYGB^{-/-}$ HepG2 lines. B) Transcription of the A_3/B exon junction (positive control) was detected in all $CYGB^{+/+}$ and $CYGB^{-/-}$ cell lines (three each, labelled A, B, and C). C) RT-PCR analysis of the B/C exon junction showed amplification only in the $CYGB^{+/+}$ cell lines. D) Schematic representation of the qRT-PCR assays used to validate the knockout of *CYGB*. E-H) qRT-PCR analysis of *CYGB* V-1 and V-3 expression in the three $CYGB^{+/+}$ and the three $CYGB^{-/-}$ HepG2 cell lines. E) Quantification of the transcription of *CYGB* V-1 and V-3 confirmed that V-3 is the predominant *CYGB* mRNA isoform in HepG2 cells. F) The expression of the 5′ end region of *CYGB* V-1 was unchanged in $CYGB^{-/-}$ cells compared to $CYGB^{+/+}$ cells (qRT-PCR assay in red). G) The expression of the 5′ end region of *CYGB* V-3 was significantly downregulated, but not absent, in $CYGB^{-/-}$ cells compared to $CYGB^{+/+}$ cells (qRT-PCR assay in green). H) A qRT-PCR assay specifically designed to amplify the deleted *CYGB* region (yellow) confirmed the successful generation of $CYGB^{-/-}$ HepG2 cell lines. $CYGB^{+/+}$: cell lines generated from single HepG2 cell clones transfected with the empty pX462 plasmid; $CYGB^{-/-}$: cell lines generated from single HepG2 cell clones transfected with gRNA-*CYGB*-pX462 plasmids; A, B, C: different $CYGB^{+/+}$ and $CYGB^{-/-}$ cell lines; NC: negative control; *: *p*-value <0.05; ***: p-value <0.001. For both genotypes $n = 3$. Error bars indicate standard errors. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

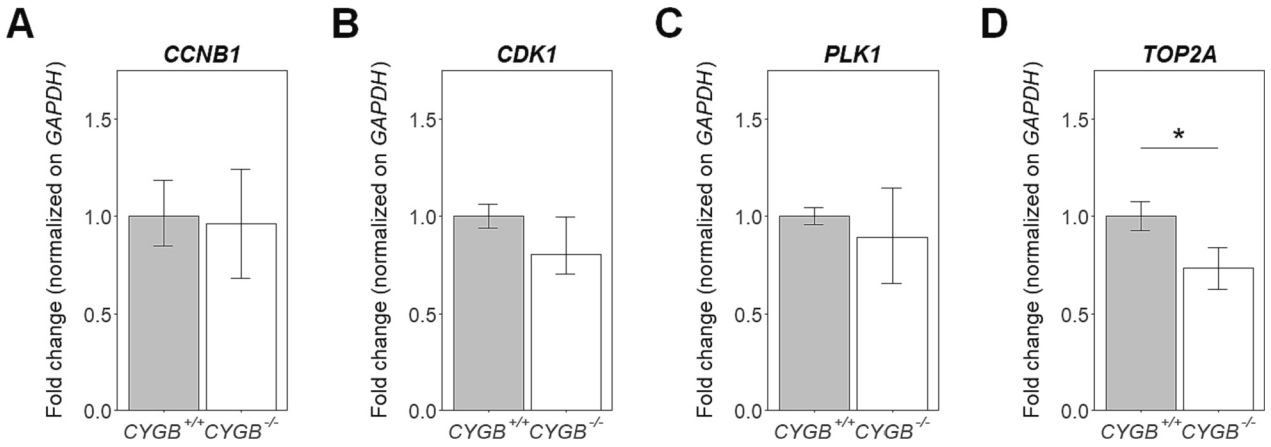


Fig. 7. Effect of *CYGB* knockout on transcription of regulators of the G2/M DNA damage checkpoint in HepG2 cells. A-D) qRT-PCR analyses showed that *CYGB* knockout did not significantly alter the transcription of *CCNB1* (A), *CDK1* (B), and *PLK1* (C) in the three tested $CYGB^{-/-}$ cell lines. *TOP2A* expression was reduced by ~30% (D). *: *p*-value <0.05; $n = 3$; Error bars indicate standard errors.

identified sequences almost perfectly match the *CYGB* mRNA isoforms annotated in the Ensembl database, but only partially fit the predicted alternative transcripts in the NCBI nucleotide database. EST and RNA-seq data confirmed that the most strongly expressed *CYGB* mRNA isoform in the majority of tissues and cells is the canonical *CYGB* V-1. While isoforms *CYGB* V-2, V-4, and V-5 lacked evidence of strong expression and may thus be of minor relevance or constitute transcriptional noise, *CYGB* V-3 was detected as prominently expressed. Notably, V-3 was the primary *CYGB* mRNA isoform in the majority of investigated normal liver and HB cancer tissue biopsies as well as in all examined HB cell lines, as confirmed at the molecular level in HepG2 cells by qRT-PCR and 5'RACE experiments. Based on these findings, we re-investigated the proportion of *CYGB* transcript variants in a recent study which focused on CYGB's role in kidney podocyte cells in the context of chronic kidney disease [39]. In fact, *CYGB* isoform V-3 was expressed in differentiated human podocyte AB8/13 cells roughly twice as much as the canonical V-1 mRNA, which underlines the necessity to take into account AS variants in prospective studies of CYGB.

To further characterize the alternative *CYGB* mRNAs and the underlying architecture of the *CYGB* gene, we performed 5'RACE in HepG2 cells and identified an exonic region ~6.4 kbp upstream of the *CYGB* locus that could be attributed to the V-3 transcript or represent the alternative first exon of an additional *CYGB* mRNA isoform, V-3r. DNA methylation, DNaseI hypersensitivity, FAIRE-seq, and Pol II ChIP-seq data from HepG2 cells (as included in the UCSC Genome browser) revealed sites of active transcription around candidate *CYGB* exons A_3 and A_{3R}, suggesting that V-3 and V-3r are two separate *CYGB* transcripts. Additional CAGE and RNA-PET data from HepG2 cells suggested that transcription from the region encompassing exon A_3 is stronger than transcription from A_{3R}. In contrast, promoter predictions identified a TATA-less candidate promoter region encompassing exon A_{3R} (GXP_4415818), while no promoter was predicted in the region surrounding exon A_3. Altogether, these analyses revealed unanticipated structural details of the human *CYGB* gene locus and a variability of transcript isoforms that has not been accounted for so far in literature.

Adding further complexity to the *CYGB* gene locus, the CAGE and RNA-PET data from the UCSC Genome browser indicated transcription of the Small Nucleolar RNA Host Gene 16 (*SNHG16* or *ncRAN*) in HepG2 cells. The first exon of this gene is located on the (+) strand of the candidate GXP_4415818 *CYGB* promoter locus (predicted on the (−) strand). Expression of *SNHG16* could potentially be explained by the presence of unmethylated CpG di-nucleotides in the CpG islands encompassing the TSS of this gene, which, because of the palindromic nature of CpGs, could regulate transcription of genes on either DNA strand [90]. *SNHG16* is a long non-coding RNA gene whose expression is associated with poor diagnosis, tumour growth and/or invasiveness in diverse cancer entities [91]. No published study has so far investigated the role of *SNHG16* in HB tissues, while three studies reported that the knockdown of *SNHG16* in HepG2 cells (incorrectly described as HCC cells in [92]) induced cell proliferation, invasion, tumorigenesis [93,94], and increased the sensitivity to sorafenib [95]. Our transcriptome analyses detected *SNHG16* expression in all investigated normal liver and HB tissue samples (Supplementary File 1), but no significant differential expression was observed between normal and tumour tissues or among biopsies from different HB subtypes (Supplementary File 2). Furthermore, only a weak correlation was found between *CYGB* and *SNHG16* transcript levels in these samples (Kendall's rank correlation $\tau = 0.20$, *p*-value = 0.04). In line with this, comprehensive transcriptome analyses indicated absence of *CYGB* mRNA in most human cancer cell lines, including lines from the majority of tumour types reported to express *SNHG16*. Together, this suggests that *CYGB* and *SNHG16* are under different transcriptional regulations and, as a consequence, that *CYGB* V-3r mRNA is not a by-product of *SNHG16* expression.

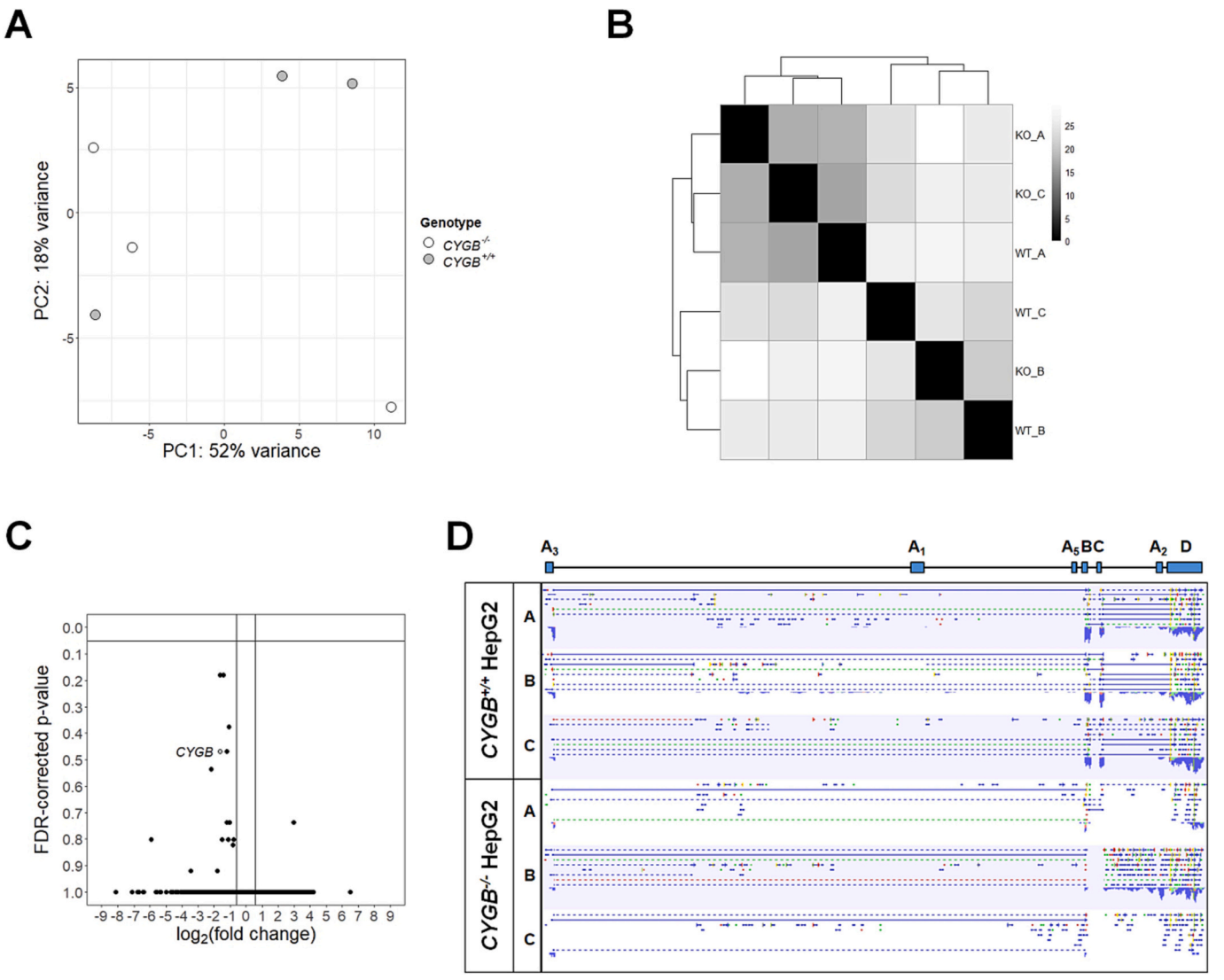


Fig. 8. *CYGB* knockout did not alter the transcriptome of HepG2 cells. A-B) Principal component analysis (A) and "sample-to-sample" distance (B) of the HepG2 datasets indicate no clear separation between $CYGB^{+/+}$ and $CYGB^{-/-}$ samples. C) Volcano plot of differential gene expression in $CYGB^{-/-}$ vs. $CYGB^{+/+}$ HepG2 samples. The analysis indicated that no gene was significantly dysregulated, including *CYGB* (gray dot). Vertical lines indicate the applied thresholds for the fold change ($\log_2$(fold change) = 0.585 and − 0.585), while the horizontal line indicates the significance threshold for the FDR-corrected p-value (0.05). D) Graphical visualization of the RNA-seq reads originated from $CYGB^{+/+}$ and $CYGB^{-/-}$ HepG2 cells that mapped to the *CYGB* locus. Transcription of the remnant *CYGB* exons in the $CYGB^{-/-}$ HepG2 lines (especially in cell line B) explains why the differential gene expression analysis did not reveal a not significant downregulation of *CYGB*. As expected, no read mapped to the deleted exonic regions, thereby confirming the editing event.

A large fraction of alternative mRNA isoforms identified by transcriptomics analyses are considered as results of transcriptional or splicing errors [96]. However, evolutionary conservation of alternative exons and splicing events can be interpreted as an indicator of biological significance [79]. We identified predicted *Cygb* transcripts with candidate *CYGB* exons A_2 or A_5 and potential homologs to *CYGB* V-3 and V-4 in several primates, which supports biological relevance of these transcript isoforms, at least in this taxon group.

4.2. Candidate CYGB protein isoforms: Hypothetical roles and limitations

Following the bioinformatical identification and experimental confirmation of alternative *CYGB* mRNA isoforms, the predictive ORF analysis showed that only isoform V-2 would not encode a functional protein. Full length CYGB would be encoded by V-1, isoforms V-3 and V5 would conceptually encode an N-terminally truncated globin lacking aa 1–65, and V-4 would encode an authentic N-terminal CYGB part of 125 aa fused to a 'novel' C-terminal region (101 aa). Notably, a short ORF potentially coding for a 36 amino aa peptide was detected upstream of the main ORF in V-5, possibly representing an upstream ORF whose transcription might inhibit translation of the primary CDS [97].

As limitation of our study, we were unable to demonstrate protein expression of the novel candidate CYGB variants due to a lack of suitable, ideally isoform-specific antibody preparations. Application of a well-established anti-rat Cygb antiserum [2], reported to recognize N-terminal aa positions 1–18 (and thus only proteins from *CYGB* V-1 and V-4), was unsuccessful in WB experiments and yielded unspecific signal. The 'novel' C-terminal part potentially encoded by V-4 could not be detected by sequence searches in mass spectrometry-based human protein expression data. Moreover, we did not succeed in expressing the truncated V-3 CYGB in standard *E. coli*-based recombinant expression systems. The proof of protein expression of the candidate CYGB variants must therefore remain a target for future research. However, we find it important to discuss the possible implications, in particular of a cellular expression of the truncated CYGB version V-3. The latter lacks the characteristic N-terminal extension of CYGB, which was suggested to

contribute to its homo-dimerization [58], and the Cys38 residue, which is crucial for *in vitro* formation of homo-dimers [56,57] and CYGB S-S monomers [57,58]. Of note, the predicted structures of the full-length and the truncated CYGB proteins reported in the AlphaFold database show almost identical 2D and 3D organizations with minor divergences in the most C-terminal parts. In full-length CYGB the formation of both the homo-dimers and the S-S monomers facilitated the binding to exogenous ligands [59,60,98], and the establishment of the intramolecular disulfide bridge induced a pentacoordinated-like state that enhances its peroxidase [34,35] and NiR activities [32]. Therefore, as the putative truncated CYGB isoform could likely neither form homo- or heterodimers nor establish an intramolecular disulfide bridges, this protein might act entirely differently from canonical CYGB.

4.3. *CYGB* expression in normal and HB liver tissues

In the context of mammalian liver tissue, the consensual *Cygb* transcript V-1 is known to be specifically expressed in HSCs [10,18,99,100], and inhibition of its expression in adult mice contributed to the establishment of various liver diseases [40,44] and malignancies [42,43]. Expression of the putative truncated CYGB isoform in this tissue, however, might have been overlooked, as many published studies investigating CYGB expression were performed using antibodies targeting CYGB's N-terminus (missing in the truncated isoform) [2]. Since *CYGB* expression is often silenced in cancer tissues and cells [14,15,101], the observation of *CYGB* transcription in HB samples and cell lines was entirely unexpected. HB is a rare neoplasm, but, the most common childhood liver cancer [102]. HB originates from immature liver precursor cells, most likely pluripotent stem cells, that can metastasize and often present mutations and/or expression alteration of genes involved in the β-catenin/Wnt signalling pathway resulting in its activation [103]. It is still unclear whether deregulation of this pathway is causally involved in HB tumour genesis or if it is only a pathologic feature of prognostic significance [71]. Our transcriptome analysis detected significantly higher *CYGB* transcription levels in C1 samples compared to C2 tumours and normal liver tissue (1.9- and 1.4- fold, respectively). Although we found V-3 to be the predominant (>94%) *CYGB* mRNA isoform in all analysed HB cancer cell lines, V-3 was co-expressed with other *CYGB* isoforms in HB biopsies and its relative expression largely varied among samples. This fluctuating expression of the different *CYGB* isoforms in HB biopsies could possibly be explained by cancer heterogeneity and/or by a different proportion of HSCs within the samples. Future studies must clarify the relative expression contribution of *CYGB* isoforms in a larger number of HB samples, accompanied by histological analysis and determination of the cellular composition.

4.4. *Knockout of CYGB in HepG2 cells: No phenotype in transcriptome and cell cycle*

The detection of appreciable levels of *CYGB* V-3 transcripts in normal and HB liver tissues as well as in multiple HB cell lines advocated for a possible functional role of the truncated CYGB variant in these tissues and cells. To investigate this, we used CRISPR/Cas9n technology to generate a genetic *CYGB* knockout in the HB cell model HepG2 by targeting the genomic sequence encoding the central domain of the protein, which contains the functionally important His81 and His113 residues. The knockout was confirmed at the genomic and transcriptional levels in the generated cells. We then performed RNA-seq on $CYGB^{+/+}$ and $CYGB^{-/-}$ HepG2 cell lines aiming at identifying genes whose expression was possibly altered by the depletion of *CYGB* V-3. Unexpectedly, differential gene expression analysis did not identify any significantly dysregulated genes. Mapping of cDNA sequence reads to the *CYGB* reference gene structure itself was limited to the remnant *CYGB* exons and the intronic region between *CYGB* exons C and D (Fig. 8D), thereby confirming successful Crispr/Cas9-mediated targeting of the gene.

Several *in vitro* studies, performed in bone osteosarcoma, ovarian cancer, and pancreatic cancer cell lines, recently proposed involvement of canonical CYGB protein in regulation of the cell cycle [51,104–106]. Moreover, our re-analysis of differentiated AB8/13 cells transcriptome data [39] revealed that a knockdown of *CYGB* induced downregulation of few regulators of the G2/M damage checkpoint (*CCNB1*, *CDK1*, *PLK1*, and *TOP2A*). Since this would suggest that *CYGB* V-3 might contribute to cell cycle regulation, we also studied this aspect in our $CYGB^{+/+}$ and $CYGB^{-/-}$ HepG2 cells. However, differently to the results in AB8/13 cells, *CYGB* knockout in HepG2 did not significantly alter expression of *CCNB1*, *CDK1*, and *PLK1*. In addition, cell cycle status analysis by flow cytometry performed on $CYGB^{+/+}$ and $CYGB^{-/-}$ HepG2 cell lines did not reveal a phenotypic effect of the knockout. Our data in the HepG2 cell model thus indicate that expression of *CYGB* V-3 may have no phenotypic effect on both the transcriptome and the cell cycle. A possible functional role of *CYGB* V-3 in HepG2 cells must therefore remain speculative.

5. Concluding remarks

This is the first study showing that the human *CYGB* gene locus has a more complicated architecture than previously described, presenting multiple transcript isoforms and novel upstream exons and candidate promoter regions. While the biological significance of the *CYGB* mRNA variants is a target for future research, the prominent expression and evolutionary conservation argues for selectable function(s), at least for transcript V-3 and its encoded truncated CYGB version. Future studies on CYGB, its function and gene regulation should account for the potential presence of these alternative transcripts.

Funding

Elena Porto received a PhD fellowship in the context of the International PhD Program funded by the Boehringer Ingelheim Foundation and the Institute of Molecular Biology (IMB) Mainz.

CRediT authorship contribution statement

Elena Porto: Conceptualization, Methodology, Investigation, Formal analysis, Data curation, Visualization, Writing – original draft. **Paraskevi Loula:** Investigation, Data curation. **Susanne Strand:** Investigation, Writing – review & editing. **Thomas Hankeln:** Conceptualization, Methodology, Supervision, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available upon request.

Acknowledgements

The authors would like to thank Prof. Dr. Norifumi Kawada (Department of Hepatology, Graduate School of Medicine, Osaka Metropolitan University, Japan) for kindly providing the anti-Cygb antibody, Prof. Dr. David Hoogewijs and Dr. Elisa Randi (University of Fribourg, Switzerland) for kindly providing HepG2, AB8/13, and AB8/13-OE cells, and Prof. Dr. Sylvia Dewilde and her group (Department of Biomedical Sciences, University of Antwerp, Belgium) for performing recombinant protein expression experiments.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jinorgbio.2023.112422.

References

[1] T. Burmester, B. Ebner, B. Weich, T. Hankeln, Cytoglobin: a novel globin type ubiquitously expressed in vertebrate tissues, Mol. Biol. Evol. 19 (2002) 416–421, https://doi.org/10.1093/oxfordjournals.molbev.a004096.

[2] N. Kawada, D.B. Kristensen, K. Asahina, K. Nakatani, Y. Minamiyama, S. Seki, K. Yoshizato, Characterization of a stellate cell activation-associated protein (STAP) with peroxidase activity found in rat hepatic stellate cells, J. Biol. Chem. 276 (2001) 25318–25323, https://doi.org/10.1074/jbc.M102630200.

[3] R.E. Dickerson, I. Geis, Hemoglobin: Structure, Function, Evolution, and Pathology, Benjamin/Cummings Publishing Company, Menlo Park, CA, 1983, https://doi.org/10.1016/0968-0004(84)90080-X.

[4] J.B. Wittenberg, B.A. Wittenberg, Myoglobin function reassessed, J. Exp. Biol. 206 (2003) 2011–2020, https://doi.org/10.1242/jeb.00243.

[5] T. Burmester, B. Weich, S. Reinhardt, T. Hankeln, A vertebrate globin expressed in the brain, Nature. 407 (2000) 520–523, https://doi.org/10.1038/35035093.

[6] D. Hoogewijs, B. Ebner, F. Germani, F.G. Hoffmann, A. Fabrizius, L. Moens, T. Burmester, S. Dewilde, J.F. Storz, S.N. Vinogradov, T. Hankeln, Androglobin: a chimeric globin in metazoans that is preferentially expressed in mammalian testes, Mol. Biol. Evol. 29 (2012) 1105–1114, https://doi.org/10.1093/molbev/msr246.

[7] T.W. Koay, C. Osterhof, I.M.C. Orlando, A. Keppner, D. Andre, S. Yousefian, M. S. Alonso, M. Correia, R. Markworth, J. Schödel, T. Hankeln, D. Hoogewijs, Androglobin gene expression patterns and FOXJ1-dependent regulation indicate its functional association with ciliogenesis, J. Biol. Chem. (2021), 100291, https://doi.org/10.1016/j.jbc.2021.100291.

[8] J.T. Trent, M.S. Hargrove, A ubiquitously expressed human hexacoordinate hemoglobin, J. Biol. Chem. 277 (2002) 19538–19545, https://doi.org/10.1074/jbc.M201934200.

[9] E. Geuens, I. Brouns, D. Flamez, S. Dewilde, J.P. Timmermans, L. Moens, A globin in the nucleus!, J. Biol. Chem. 278 (2003) 30417–30420, https://doi.org/10.1074/jbc.C300203200.

[10] K. Nakatani, H. Okuyama, Y. Shimahara, S. Saeki, D.-H. Kim, Y. Nakajima, S. Seki, N. Kawada, K. Yoshizato, Cytoglobin/STAP, its unique localization in splanchnic fibroblast-like cells and function in organ fibrogenesis, Lab. Investig. 84 (2004) 91–101, https://doi.org/10.1038/sj.labinvest.3700013.

[11] M. Schmidt, F. Gerlach, A. Avivi, T. Laufs, S. Wystub, J.C. Simpson, E. Nevo, S. Saaler-Reinhardt, S. Reuss, T. Hankeln, T. Burmester, Cytoglobin is a respiratory protein in connective tissue and neurons, which is up-regulated by hypoxia, J. Biol. Chem. 279 (2004) 8063–8069, https://doi.org/10.1074/jbc.M310540200.

[12] M. Schmidt, T. Laufs, S. Reuss, T. Hankeln, T. Burmester, Divergent distribution of cytoglobin and neuroglobin in the murine eye, Neurosci. Lett. 374 (2005) 207–211, https://doi.org/10.1016/j.neulet.2004.10.071.

[13] A. Shigematsu, Y. Adachi, J. Matsubara, H. Mukaide, N. Koike-Kiriyama, K. Minamino, M. Shi, S. Yanai, M. Imamura, S. Taketani, S. Ikehara, Analyses of expression of Cytoglobin by Immunohistochemical studies in human tissues, Hemoglobin. 32 (2008) 287–296, https://doi.org/10.1080/03630260802017261.

[14] T.A. Gorr, D. Wichmann, C. Pilarsky, J.P. Theurillat, A. Fabrizius, T. Laufs, T. Bauer, M. Koslowski, S. Horn, T. Burmester, T. Hankeln, G. Kristiansen, Old proteins - new locations: myoglobin, haemoglobin, neuroglobin and cytoglobin in solid tumours and cancer cells, Acta Physiol (Oxford) 202 (2011) 563–581, https://doi.org/10.1111/j.1748-1716.2010.02205.x.

[15] Y. Fujita, S. Koinuma, M.A. De Velasco, J. Bolz, Y. Togashi, M. Terashima, H. Hayashi, T. Matsuo, K. Nishio, Melanoma transition is frequently accompanied by a loss of cytoglobin expression in melanocytes: a novel expression site of cytoglobin, PLoS One 9 (2014) 1–9, https://doi.org/10.1371/journal.pone.0094772.

[16] A. Latina, G. Viticchiè, A.M. Lena, M.C. Piro, M. Annicchiarico-Petruzzelli, G. Melino, E. Candi, ΔNp63 targets cytoglobin to inhibit oxidative stress-induced apoptosis in keratinocytes and lung cancer, Oncogene (2015) 1–11, https://doi.org/10.1038/onc.2015.222.

[17] A. Doğan, S. Demirci, B. Kıratlı, F. Şahin, Cytoglobin: a potential marker for adipogenic differentiation in preadipocytes in vitro, Cytotechnology. 69 (2017) 157–165, https://doi.org/10.1007/s10616-016-0047-2.

[18] T. Hankeln, S. Wystub, T. Laufs, M. Schmidt, F. Gerlach, S. Saaler-Reinhardt, S. Reuss, T. Burmester, The cellular and subcellular localization of neuroglobin and cytoglobin – a clue to their function? IUBMB Life 56 (2004) 671–679, https://doi.org/10.1080/15216540500037794.

[19] S. Singh, S.M. Manda, D. Sikder, M.J. Birrer, B.A. Rothermel, D.J. Garry, P.P. A. Mammen, Calcineurin activates cytoglobin transcription in hypoxic myocytes, J. Biol. Chem. 284 (2009) 10409–10421, https://doi.org/10.1074/jbc.M809572200.

[20] H. Li, C. Hemann, T.M. Abdelghany, M.A. El-Mahdy, J.L. Zweier, Characterization of the mechanism and magnitude of cytoglobin-mediated nitrite reduction and nitric oxide generation under anaerobic conditions, J. Biol. Chem. 287 (2012) 36623–36633, https://doi.org/10.1074/jbc.M112.342378.

[21] F.L. Jourd'heuil, H. Xu, T. Reilly, K. McKellar, C. El Alaoui, J. Steppich, Y.F. Liu, W. Zhao, R. Ginnan, D. Conti, R. Lopez-Soler, A. Asif, R.K. Keller, J.J. Schwarz, L. T. Thanh Thuy, N. Kawada, X. Long, H.A. Singer, D. Jourd'heuil, The hemoglobin homolog Cytoglobin in smooth muscle inhibits apoptosis and regulates vascular RemodelingHighlights, Arterioscler. Thromb. Vasc. Biol. 37 (2017) 1944–1955, https://doi.org/10.1161/ATVBAHA.117.309410.

[22] B. Lilly, K. Dammeyer, S. Marosis, P.E. McCallinhart, A.J. Trask, M. Lowe, D. Sawant, Endothelial cell-induced cytoglobin expression in vascular smooth muscle cells contributes to modulation of nitric oxide, Vasc. Pharmacol. 110 (2018) 7–15, https://doi.org/10.1016/j.vph.2018.06.016.

[23] X. Liu, M.A. El-Mahdy, J. Boslett, S. Varadharaj, C. Hemann, T.M. Abdelghany, R. S. Ismail, S.C. Little, D. Zhou, L.T.T. Thuy, N. Kawada, J.L. Zweier, Cytoglobin regulates blood pressure and vascular tone through nitric oxide metabolism in the vascular wall, Nat. Commun. 8 (2017) 14807, https://doi.org/10.1038/ncomms14807.

[24] G. Ilangovan, S.A. Khaleel, T. Kundu, C. Hemann, M.A. El-Mahdy, J.L. Zweier, Defining the reducing system of the NO dioxygenase cytoglobin in vascular smooth muscle cells and its critical role in regulating cellular NO decay, J. Biol. Chem. 296 (2021), 100196, https://doi.org/10.1074/jbc.ra120.016394.

[25] N. Presneau, K. Dewar, V. Forgetta, D. Provencher, A.M. Mes-Masson, P.N. Tonin, Loss of heterozygosity and transcriptome analyses of a 1.2 Mb candidate ovarian cancer tumor suppressor locus region at 17q25.1-q25.2, Mol. Carcinog. 43 (2005) 141–154, https://doi.org/10.1002/mc.20096.

[26] B.J. Smagghe, G. Sarath, E. Ross, J.L. Hilbert, M.S. Hargrove, Slow ligand binding kinetics dominate ferrous hexacoordinate hemoglobin reactivities and reveal differences between plants and other species, Biochemistry. 45 (2006) 561–570, https://doi.org/10.1021/bi051902l.

[27] K.E. Halligan, F.L. Jourd'heuil, D. Jourd'heuil, Cytoglobin is expressed in the vasculature and regulates cell respiration and proliferation via nitric oxide dioxygenation, J. Biol. Chem. 284 (2009) 8539–8547, https://doi.org/10.1074/jbc.M808231200.

[28] A.M. Gardner, M.R. Cook, P.R. Gardner, Nitric-oxide dioxygenase function of human cytoglobin with cellular reductants and in rat hepatocytes, J. Biol. Chem. 285 (2010) 23850–23857, https://doi.org/10.1074/jbc.M110.132340.

[29] X. Liu, J. Tong, J.R. Zweier, D. Follmer, C. Hemann, R.S. Ismail, J.L. Zweier, Differences in oxygen-dependent nitric oxide metabolism by cytoglobin and myoglobin account for their differing functional roles, FEBS J. 280 (2013) 3621–3631, https://doi.org/10.1111/febs.12352.

[30] J. Tong, J.R. Zweier, R.L. Huskey, R.S. Ismail, C. Hemann, J.L. Zweier, X. Liu, Effect of temperature, pH and heme ligands on the reduction of Cygb(Fe 3+) by ascorbate, Arch. Biochem. Biophys. 554 (2014) 1–5, https://doi.org/10.1016/j.abb.2014.04.011.

[31] M.B. Amdahl, C.E. Sparacino-Watkins, P. Corti, M.T. Gladwin, J. Tejero, Efficient reduction of vertebrate Cytoglobins by the cytochrome b5/cytochrome b5Reductase/NADH system, Biochemistry. 56 (2017) 3993–4004, https://doi.org/10.1021/acs.biochem.7b00224.

[32] B.J. Reeder, J. Ukeri, Strong modulation of nitrite reductase activity of cytoglobin by disulfide bond oxidation: implications for nitric oxide homeostasis, Nitric Oxide Biol. Chem. 72 (2018) 16–23, https://doi.org/10.1016/j.niox.2017.11.004.

[33] J.L. Zweier, C. Hemann, T. Kundu, M.G. Ewees, S.A. Khaleel, A. Samouilov, G. Ilangovan, M.A. El-Mahdy, Cytoglobin has potent superoxide dismutase function, Proc. Natl. Acad. Sci. U. S. A. 118 (2021) 1–12, https://doi.org/10.1073/pnas.2105053118.

[34] P. Beckerson, D. Svistunenko, B. Reeder, Effect of the distal histidine on the peroxidatic activity of monomeric cytoglobin, F1000Research. 4 (2015) 87, https://doi.org/10.12688/f1000research.5971.1.

[35] J. Tejero, A.A. Kapralov, M.P. Baumgartner, C.E. Sparacino-Watkins, T. S. Anthonymutu, I.I. Vlasova, C.J. Camacho, M.T. Gladwin, H. Bayir, V.E. Kagan, Peroxidase activation of cytoglobin by anionic phospholipids: mechanisms and consequences, Biochim. Biophys. Acta Mol. Cell Biol. Lipids 1861 (2016) 391–401, https://doi.org/10.1016/j.bbalip.2016.02.022.

[36] B.J. Reeder, D.A. Svistunenko, M.T. Wilson, Lipid binding to cytoglobin leads to a change in haem co-ordination: a role for cytoglobin in lipid signalling of oxidative stress, Biochem. J. 434 (2011) 483–492, https://doi.org/10.1042/BJ20101136.

[37] S. Singh, D.C. Canseco, S.M. Manda, J.M. Shelton, R.R. Chirumamilla, S. C. Goetsch, Q. Ye, R.D. Gerard, J.W. Schneider, J.A. Richardson, B.A. Rothermel, P.P.A. Mammen, Cytoglobin modulates myogenic progenitor cell viability and muscle regeneration, Proc. Natl. Acad. Sci. U. S. A. 111 (2014) E129–E138, https://doi.org/10.1073/pnas.1314962111.

[38] N.J. Hodges, N. Innocent, S. Dhanda, M. Graham, Cellular protection from oxidative DNA damage by over-expression of the novel globin cytoglobin in vitro, Mutagenesis. 23 (2008) 293–298, https://doi.org/10.1093/mutage/gen013.

[39] E.B. Randi, B. Vervaet, M. Tsachaki, E. Porto, S. Vermeylen, M.T. Lindenmeyer, L. T.T. Thuy, C. Cohen, O. Devuyst, A.D. Kistler, C. Szabo, N. Kawada, T. Hankeln, A. Odermatt, S. Dewilde, R.H. Wenger, D. Hoogewijs, The anti-oxidative role of cytoglobin in podocytes: implications for a role in chronic kidney disease, Antioxid. Redox Signal. 3 (2020) 1–55, https://doi.org/10.1089/ars.2019.7868.

[40] T.T. Van Thuy, L.T.T. Thuy, K. Yoshizato, N. Kawada, G.M. Hirschfield, E. J. Heathcote, M.E. Gershwin, M.A. Aller, C. Rust, B.L. Copple, H. Jaeschke, C. D. Klaassen, N. Kawada, K. Nakatani, H. Sawai, A.M. Gardner, M.R. Cook, P. R. Gardner, X. Liu, M. Emara, A.R. Turner, J. Allalunis-Turner, B.J. Smagghe, J. T. Trent, M.S. Hargrove, X. Liu, P. Pacher, J.S. Beckman, L. Liaudet, T.T.T. Le, T. T.T. Le, T.T.T. Le, Y. Fang, A.P. Rolo, C.M. Palmeira, J.M. Holy, K.B. Wallace, Y. Tsujimoto, S. Shimizu, H. Miyoshi, C. Rust, M.E. Guicciardi, G.J. Gores, J. Lee, Y. Takakuwa, S.L. Loke, H. Sugimoto, M.G. Clemens, H. Jaeschke, Y.M. Kim, R.

V. Talanian, T.R. Billiar, J.F. Dufour, T.J. Turner, I.M. Arias, M.Y. Balakirev, V. V. Khramtsov, G. Zimmer, C.M. Schonhoff, U. Ramasamy, M.S. Anwer, G.L. Tipoe, H. Shafaroodi, P. Mayoral, K. Allen, H. Jaeschke, B.L. Copple, J. Weerachayaphorn, A.R. Soylu, H. Yang, I. Mimura, Possible involvement of nitric oxide in enhanced liver injury and Fibrogenesis during cholestasis in Cytoglobin-deficient mice, Sci. Rep. 7 (2017) 41888, https://doi.org/10.1038/srep41888.

[41] Y. Okina, M. Sato-Matsubara, Y. Kido, H. Urushima, A. Daikoku, C. Kadono, Y. Nakagama, Y. Nitahara, T.H. Hoang, L.T.T. Thuy, T. Matsubara, N. Ohtani, K. Ikeda, K. Yoshizato, N. Kawada, Nitric oxide derived from cytoglobin-deficient hepatic stellate cells causes suppression of cytochrome c oxidase activity in hepatocytes, Antioxid. Redox Signal. (2023), https://doi.org/10.1089/ars.2021.0279.

[42] L. Thi Thanh Thuy, T. Morita, K. Yoshida, K. Wakasa, M. Iizuka, T. Ogawa, M. Mori, Y. Sekiya, S. Momen, H. Motoyama, K. Ikeda, K. Yoshizato, N. Kawada, Promotion of liver and lung tumorigenesis in DEN-treated cytoglobin-deficient mice, Am. J. Pathol. 179 (2011) 1050–1060, https://doi.org/10.1016/j.ajpath.2011.05.006.

[43] L.T.T. Thuy, Y. Matsumoto, T.T. Van Thuy, H. Hai, M. Suoh, Y. Urahara, H. Motoyama, H. Fujii, A. Tamori, S. Kubo, S. Takemura, T. Morita, K. Yoshizato, N. Kawada, Cytoglobin deficiency promotes liver Cancer development from Hepatosteatosis through activation of the oxidative stress pathway, Am. J. Pathol. 185 (2015) 1045–1060, https://doi.org/10.1016/j.ajpath.2014.12.017.

[44] L.T.T. Thuy, T.T. Van Thuy, Y. Matsumoto, H. Hai, Y. Ikura, K. Yoshizato, N. Kawada, Absence of cytoglobin promotes multiple organ abnormalities in aged mice, Sci. Rep. 6 (2016) 24990, https://doi.org/10.1038/srep24990.

[45] M. Yassin, H. Kissow, B. Vainer, P.D. Joseph, A. Hay-Schmidt, J. Olsen, A. E. Pedersen, Cytoglobin affects tumorigenesis and the expression of ulcerative colitis-associated genes under chemically induced colitis in mice, Sci. Rep. 8 (2018) 6905, https://doi.org/10.1038/s41598-018-24728-x.

[46] R.J. Shaw, M.M. Omar, S. Rokadiya, F.A. Kogera, D. Lowe, G.L. Hall, J. A. Woolgar, J. Homer, T. Liloglou, J.K. Field, J.M. Risk, Cytoglobin is upregulated by tumour hypoxia and silenced by promoter hypermethylation in head and neck cancer, Br. J. Cancer 101 (2009) 139–144, https://doi.org/10.1038/sj.bjc.6605121.

[47] M. Emara, A.R. Turner, J. Allalunis-Turner, Hypoxic regulation of cytoglobin and neuroglobin expression in human normal and tumor tissues, Cancer Cell Int. 10 (2010) 33, https://doi.org/10.1186/1475-2867-10-33.

[48] S. Wystub, B. Ebner, C. Fuchs, B. Weich, T. Burmester, T. Hankeln, Interspecies comparison of neuroglobin, cytoglobin and myoglobin: sequence evolution and candidate regulatory elements, Cytogenet. Genome Res. 105 (2004) 65–78, https://doi.org/10.1159/000078011.

[49] H.-W. Xu, Y.-J. Huang, Z.-Y. Xie, L. Lin, Y.-C. Guo, Z.-R. Zhuang, X.-P. Lin, W. Zhou, M. Li, H.-H. Huang, X.-L. Wei, K. Man, G.-J. Zhang, The expression of cytoglobin as a prognostic factor in gliomas: a retrospective analysis of 88 patients, BMC Cancer 13 (2013) 247, https://doi.org/10.1186/1471-2407-13-247.

[50] P.M. Wojnarowicz, D.M. Provencher, A.M. Mes-Masson, P.N. Tonin, Chromosome 17q25 genes, RHBDF2 and CYGB, inovarian cancer, Int. J. Oncol. 40 (2012) 1865–1880, https://doi.org/10.3892/ijo.2012.1371.

[51] H. Chen, X. Zhao, T. Meng, Expression and biological role of cytoglobin in human ovarian cancer, Tumor Biol. 35 (2014) 6933–6939, https://doi.org/10.1007/s13277-014-1941-x.

[52] Y. Feng, M. Wu, S. Li, X. He, J. Tang, W. Peng, B. Zeng, C. Deng, G. Ren, T. Xiang, The epigenetically downregulated factor CYGB suppresses breast cancer through inhibition of glucose metabolism, J. Exp. Clin. Cancer Res. 37 (2018) 313, https://doi.org/10.1186/s13046-018-0979-9.

[53] X. Guo, S. Philipsen, K.C. Tan-Un, Study of the hypoxia-dependent regulation of human CYGB gene, Biochem. Biophys. Res. Commun. 364 (2007) 145–150, https://doi.org/10.1016/j.bbrc.2007.09.108.

[54] T. Hankeln, B. Ebner, C. Fuchs, F. Gerlach, M. Haberkamp, T.L. Laufs, A. Roesner, M. Schmidt, B. Weich, S. Wystub, S. Saaler-Reinhardt, S. Reuss, M. Bolognesi, D. De Sanctis, M.C. Marden, L. Kiger, L. Moens, S. Dewilde, E. Nevo, A. Avivi, R. E. Weber, A. Fago, T. Burmester, Neuroglobin and cytoglobin in search of their role in the vertebrate globin family, J. Inorg. Biochem. 99 (2005) 110–119, https://doi.org/10.1016/j.jinorgbio.2004.11.009.

[55] S. Hanai, H. Tsujino, T. Yamashita, R. Torii, H. Sawai, Y. Shiro, K. Oohora, T. Hayashi, T. Uno, Roles of N- and C-terminal domains in the ligand-binding properties of cytoglobin, J. Inorg. Biochem. 179 (2018) 1–9, https://doi.org/10.1016/j.jinorgbio.2017.11.003.

[56] H. Sugimoto, M. Makino, H. Sawai, N. Kawada, K. Yoshizato, Y. Shiro, Structural basis of human cytoglobin for ligand binding, J. Mol. Biol. 339 (2004) 873–885, https://doi.org/10.1016/j.jmb.2004.04.024.

[57] C. Lechauve, C. Chauvierre, S. Dewilde, L. Moens, B.N. Green, M.C. Marden, C. Clier, L. Kiger, Cytoglobin conformations and disulfide bond formation, FEBS J. 277 (2010) 2696–2704, https://doi.org/10.1111/j.1742-464X.2010.07686.x.

[58] D. Hamdane, L. Kiger, S. Dewilde, B.N. Green, A. Pesce, J. Uzan, T. Burmester, T. Hankeln, M. Bolognesi, L. Moens, M.C. Marden, The redox state of the cell regulates the ligand binding affinity of human Neuroglobin and Cytoglobin, J. Biol. Chem. 278 (2003) 51713–51721, https://doi.org/10.1074/jbc.M309396200.

[59] P. Beckerson, M.T. Wilson, D.A. Svistunenko, B.J. Reeder, Cytoglobin ligand binding regulated by changing haem-co-ordination in response to intramolecular disulfide bond formation and lipid interaction, Biochem. J. 465 (2015) 127–137, https://doi.org/10.1042/BJ20140827.

[60] P. Beckerson, B.J. Reeder, M.T. Wilson, Coupling of disulfide bond and distal histidine dissociation in human ferrous cytoglobin regulates ligand binding, FEBS Lett. 589 (2015) 507–512, https://doi.org/10.1016/j.febslet.2015.01.010.

[61] D. Zhou, C. Hemann, J. Boslett, A. Luo, J.L. Zweier, X. Liu, Oxygen binding and nitric oxide dioxygenase activity of cytoglobin are altered to different extents by cysteine modification, FEBS Open Bio 7 (2017) 845–853, https://doi.org/10.1002/2211-5463.12230.

[62] Q. Pan, O. Shai, L.J. Lee, B.J. Frey, B.J. Blencowe, Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing, Nat. Genet. 40 (2008) 1413–1415, https://doi.org/10.1038/ng.259.

[63] E.T. Wang, R. Sandberg, S. Luo, I. Khrebtukova, L. Zhang, C. Mayr, S. F. Kingsmore, G.P. Schroth, C.B. Burge, Alternative isoform regulation in human tissue transcriptomes, Nature. 456 (2008) 470–476, https://doi.org/10.1038/nature07509.

[64] D.L. Wheeler, D.M. Church, S. Federhen, A.E. Lash, T.L. Madden, J.U. Pontius, G. D. Schuler, L.M. Schriml, E. Sequeira, T.A. Tatusova, L. Wagner, Database resources of the national center for biotechnology, Nucleic Acids Res. 31 (2003) 28–33, https://doi.org/10.1093/nar/gkg033.

[65] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A.A. Kohl, A.J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A.W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, Highly accurate protein structure prediction with AlphaFold, Nature. 596 (2021) 583–589, https://doi.org/10.1038/s41586-021-03819-2.

[66] M. Varadi, S. Anyango, M. Deshpande, S. Nair, C. Natassia, G. Yordanova, D. Yuan, O. Stroe, G. Wood, A. Laydon, A. Žídek, T. Green, K. Tunyasuvunakool, S. Petersen, J. Jumper, E. Clancy, R. Green, A. Vora, M. Lutfi, M. Figurnov, A. Cowie, N. Hobbs, P. Kohli, G. Kleywegt, E. Birney, D. Hassabis, S. Velankar, AlphaFold protein structure database: massively expanding the structural coverage of protein-sequence space with high-accuracy models, Nucleic Acids Res. 50 (2022) D439–D444, https://doi.org/10.1093/nar/gkab1061.

[67] S.K. Burley, C. Bhikadiya, C. Bi, S. Bittrich, H. Chao, L. Chen, P.A. Craig, G. V. Crichlow, K. Dalenberg, J.M. Duarte, S. Dutta, M. Fayazi, Z. Feng, J.W. Flatt, S. Ganesan, S. Ghosh, D.S. Goodsell, R.K. Green, V. Guranovic, J. Henry, B. P. Hudson, I. Khokhriakov, C.L. Lawson, Y. Liang, R. Lowe, E. Peisach, I. Persikova, D.W. Piehl, Y. Rose, A. Sali, J. Segura, M. Sekharan, C. Shao, B. Vallat, M. Voigt, B. Webb, J.D. Westbrook, S. Whetstone, J.Y. Young, A. Zalevsky, C. Zardecki, RCSB protein data Bank (RCSB.Org): delivery of experimentally-determined PDB structures alongside one million computed structure models of proteins from artificial intelligence/machine learning, Nucleic Acids Res. 51 (2023) D488–D508, https://doi.org/10.1093/nar/gkac1077.

[68] K.B. Hooks, J. Audoux, H. Fazli, S. Lesjean, T. Ernault, N.-D. Senant, T. Leste-Lasserre, M. Hagedorn, B. Rousseau, C. Danet, S. Branchereau, L. Brugières, S. Taque, C. Guettier, M. Fabre, A. Rullier, M.-A. Buendia, T. Commes, C. F. Grosset, A.-A. Raymond, New insights into diagnosis and therapeutic options for proliferative hepatoblastoma, Hepatology (2017) 1–36, https://doi.org/10.1002/hep.29672.

[69] S. Cairo, C. Armengol, A. De Reyniès, Y. Wei, E. Thomas, C.A. Renard, A. Goga, A. Balakrishnan, M. Semeraro, L. Gresh, M. Pontoglio, H. Strick-Marchand, F. Levillayer, Y. Nouet, D. Rickman, F. Gauthier, S. Branchereau, L. Brugières, V. Laithier, R. Bouvier, F. Boman, G. Basso, J.F. Michiels, P. Hofman, F. Arbez-Gindre, H. Jouan, M.C. Rousselet-Chapeau, D. Berrebi, L. Marcellin, F. Plenat, D. Zachar, M. Joubert, J. Selves, D. Pasquier, P. Bioulac-Sage, M. Grotzer, M. Childs, M. Fabre, M.A. Buendia, Hepatic stem-like phenotype and interplay of Wnt/β-catenin and Myc signaling in aggressive childhood liver Cancer, Cancer Cell 14 (2008) 471–484, https://doi.org/10.1016/j.ccr.2008.11.002.

[70] D. Bell, S. Ranganathan, J. Tao, S.P. Monga, Novel advances in understanding of molecular pathogenesis of Hepatoblastoma: a Wnt/β-catenin perspective, Gene Expr. 17 (2017) 141–154, https://doi.org/10.3727/105221616X693639.

[71] P. Sumazin, Y. Chen, L.R. Treviño, S.F. Sarabia, O.A. Hampton, K. Patel, T. A. Mistretta, B. Zorman, P. Thompson, A. Heczey, S. Comerford, D.A. Wheeler, M. Chintagumpala, R. Meyers, D. Rakheja, M.J. Finegold, G. Tomlinson, D. W. Parsons, D. López-Terrada, Genomic analysis of hepatoblastoma identifies distinct molecular and prognostic subgroups, Hepatology. 65 (2017) 104–121, https://doi.org/10.1002/hep.28888.

[72] A. Mortazavi, B.A. Williams, K. McCue, L. Schaeffer, B. Wold, Mapping and quantifying mammalian transcriptomes by RNA-Seq, Nat. Methods 5 (2008) 621–628, https://doi.org/10.1038/nmeth.1226.

[73] M.I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2, Genome Biol. 15 (2014) 550, https://doi.org/10.1186/s13059-014-0550-8.

[74] V.V. Solovyev, I.A. Shahmuradov, A.A. Salamov, Identification of Promoter Regions and Regulatory Sites, 2010, pp. 57–83, https://doi.org/10.1007/978-1-60761-854-6_5.

[75] T.Y. Lee, W.C. Chang, J.B.K. Hsu, T.H. Chang, D.M. Shien, GPMiner: an integrated system for mining combinatorial cis-regulatory elements in mammalian gene group, Ser. Adv. Bioinforma. Comput. Biol. 13 (2012) 1–12, https://doi.org/10.1186/1471-2164-13-S1-S3.

[76] M.M. Konarska, P.J. Grabowski, R.A. Padgett, P.A. Sharp, Characterization of the branch site in lariat RNAs produced by splicing of mRNA precursors, Nature. 313 (1985) 552–557, https://doi.org/10.1038/313552a0.

[77] J.V. Dacie, N.K. Shinton, P.J. Gaffney, R.W. Carrell, H. Lehmann, Haemoglobin Hammersmith (β 42 (CD1) Phe→Ser), Nature. 216 (1967) 663–665, https://doi.org/10.1038/216663a0.
[78] M. Ota, Y. Isogai, K. Nishikawa, Structural requirement of highly-conserved residues in globins, FEBS Lett. 415 (1997) 129–133, https://doi.org/10.1016/S0014-5793(97)01110-1.
[79] M. Kozak, Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes, Cell. 44 (1986) 283–292, https://doi.org/10.1016/0092-8674(86)90762-2.
[80] M. Kozak, Recognition of AUG and alternative initiator codons is augmented by G in position +4 but is not generally affected by the nucleotides in positions +5 and +6, EMBO J. 16 (1997) 2482–2492, https://doi.org/10.1093/emboj/16.9.2482.
[81] M. Kozak, Emerging links between initiation of translation and human diseases, Mamm. Genome 13 (2002) 401–410, https://doi.org/10.1007/s00335-002-4002-5.
[82] P. Carninci, A. Sandelin, B. Lenhard, S. Katayama, K. Shimokawa, J. Ponjavic, C. A.M. Semple, M.S. Taylor, P.G. Engström, M.C. Frith, A.R.R. Forrest, W. B. Alkema, S.L. Tan, C. Plessy, R. Kodzius, T. Ravasi, T. Kasukawa, S. Fukuda, M. Kanamori-Katayama, Y. Kitazume, H. Kawaji, C. Kai, M. Nakamura, H. Konno, K. Nakano, S. Mottagui-Tabar, P. Arner, A. Chesi, S. Gustincich, F. Persichetti, H. Suzuki, S.M. Grimmond, C.A. Wells, V. Orlando, C. Wahlestedt, E.T. Liu, M. Harbers, J. Kawai, V.B. Bajic, D.A. Hume, Y. Hayashizaki, Genome-wide analysis of mammalian promoter architecture and evolution, Nat. Genet. 38 (2006) 626–635, https://doi.org/10.1038/ng1789.
[83] R.N. Nurtdinov, A.D. Neverov, A.V. Favorov, A.A. Mironov, M.S. Gelfand, Conserved and species-specific alternative splicing in mammalian genomes, BMC Evol. Biol. 7 (2007) 1–12, https://doi.org/10.1186/1471-2148-7-249.
[84] J. Pines, T. Hunter, Isolation of a human cyclin cDNA: evidence for cyclin mRNA and protein regulation in the cell cycle and for interaction with p34cdc2, Cell. 58 (1989) 833–846, https://doi.org/10.1016/0092-8674(89)90936-7.
[85] G. Draetta, D. Beach, Activation of cdc2 protein kinase during mitosis in human cells: cell cycle-dependent phosphorylation and subunit rearrangement, Cell. 54 (1988) 17–26, https://doi.org/10.1016/0092-8674(88)90175-4.
[86] L. Gheghiani, D. Loew, B. Lombard, J. Mansfeld, O. Gavet, PLK1 activation in late G2 sets up commitment to mitosis, Cell Rep. 19 (2017) 2060–2073, https://doi.org/10.1016/J.CELREP.2017.05.031.
[87] J.J. Bower, G.F. Karaca, Y. Zhou, D.A. Simpson, M. Cordeiro-Stone, W. K. Kaufmann, Topoisomerase IIα maintains genomic stability through decatenation G2 checkpoint signaling, Oncogene. 29 (2010) 4787–4799, https://doi.org/10.1038/onc.2010.232.
[88] B.R. Graveley, Alternative splicing: increasing diversity in the proteomic world, Trends Genet. 17 (2001) 100–107, https://doi.org/10.1016/S0168-9525(00)02176-4.
[89] F. Birzele, G. Csaba, R. Zimmer, Alternative splicing and protein structure evolution, Nucleic Acids Res. 36 (2008) 550–558, https://doi.org/10.1093/nar/gkm1054.
[90] N.D. Trinklein, S. Force Aldred, S.J. Hartman, D.I. Schroeder, R.P. Otillar, R. M. Myers, An abundance of bidirectional promoters in the human genome, Genome Res. 14 (2004) 62–66, https://doi.org/10.1101/gr.1982804.
[91] S. Ghafouri-Fard, T. Khoshbakht, M. Taheri, S. Shojaei, A review on the role of small nucleolar RNA host gene 6 long non-coding RNAs in the carcinogenic processes, Front. Cell Dev. Biol. 9 (2021) 1–13, https://doi.org/10.3389/fcell.2021.741684.
[92] D. López-Terrada, S.W. Cheung, M.J. Finegold, B.B. Knowles, Hep G2 is a hepatoblastoma-derived cell line, Hum. Pathol. 40 (2009) 1512–1515, https://doi.org/10.1016/j.humpath.2009.07.003.
[93] X. Xie, X. Xu, C. Sun, Z. Yu, Long intergenic noncoding RNA SNHG16 interacts with miR-195 to promote proliferation, invasion and tumorigenesis in hepatocellular carcinoma, Exp. Cell Res. 383 (2019), 111501, https://doi.org/10.1016/j.yexcr.2019.111501.
[94] Q.J. Zhang, D.Z. Li, B.Y. Lin, L. Geng, Z. Yang, S.S. Zheng, SNHG16 promotes hepatocellular carcinoma development via activating ECM receptor interaction pathway, Hepatobiliary Pancreat. Dis. Int. 21 (2022) 41–49, https://doi.org/10.1016/j.hbpd.2021.09.006.
[95] J. Ye, R. Zhang, X. Du, W. Chai, Q. Zhou, Long noncoding RNA SNHG16 induces sorafenib resistance in hepatocellular carcinoma cells through sponging miR-140-5p, Onco Targets Ther 12 (2019) 415–422, https://doi.org/10.2147/OTT.S175176.
[96] M.L. Tress, F. Abascal, A. Valencia, Alternative splicing may not be the key to proteome complexity, Trends Biochem. Sci. 42 (2017) 98–110, https://doi.org/10.1016/j.tibs.2016.08.008.
[97] M. Kozak, Selection of initiation sites by eucaryotic ribosomes: effect of inserting AUG triplets upstream from the coding sequence for preproinsulin, Nucleic Acids Res. 12 (1984) 3873–3893, https://doi.org/10.1093/nar/12.9.3873.
[98] A. Fago, C. Hundahl, S. Dewilde, K. Gilany, L. Moens, R.E. Weber, Allosteric regulation and temperature dependence of oxygen binding in human neuroglobin and cytoglobin: molecular mechanisms and physiological significance, J. Biol. Chem. 279 (2004) 44417–44426, https://doi.org/10.1074/jbc.M407126200.
[99] H. Motoyama, T. Komiya, L.T.T. Thuy, A. Tamori, M. Enomoto, H. Morikawa, S. Iwai, S. Uchida-Kobayashi, H. Fujii, A. Hagihara, E. Kawamura, Y. Murakami, K. Yoshizato, N. Kawada, Cytoglobin is expressed in hepatic stellate cells, but not in myofibroblasts, in normal and fibrotic human liver, Lab. Investig. 94 (2014) 192–207, https://doi.org/10.1038/labinvest.2013.135.
[100] A. Avivi, F. Gerlach, A. Joel, S. Reuss, T. Burmester, E. Nevo, T. Hankeln, Neuroglobin, cytoglobin, and myoglobin contribute to hypoxia adaptation of the subterranean mole rat Spalax, Proc. Natl. Acad. Sci. 107 (2010) 21570–21575, https://doi.org/10.1073/pnas.1015379107.
[101] N. Shivapurkar, V. Stastny, N. Okumura, L. Girard, Y. Xie, C. Prinsen, F. B. Thunnissen, I.I. Wistuba, B. Czerniak, E. Frenkel, J.A. Roth, T. Liloglou, G. Xinarianos, J.K. Field, J.D. Minna, A.F. Gazdar, Cytoglobin, the newest member of the globin family, functions as a tumor suppressor gene, Cancer Res. 68 (2008) 7448–7456, https://doi.org/10.1158/0008-5472.CAN-08-0565.
[102] P. Czauderna, B. Haeberle, E. Hiyama, A. Rangaswami, M. Krailo, R. Maibach, E. Rinaldi, Y. Feng, D. Aronson, M. Malogolowkin, K. Yoshimura, I. Leuschner, D. Lopez-Terrada, T. Hishiki, G. Perilongo, D. Von Schweinitz, I. Schmid, K. Watanabe, M. Derosa, R. Meyers, The Children's hepatic tumors international collaboration (CHIC): novel global rare tumor database yields new prognostic factors in hepatoblastoma and becomes a research model, Eur. J. Cancer 52 (2016) 92–101, https://doi.org/10.1016/j.ejca.2015.09.023.
[103] A. Koch, Elevated expression of Wnt antagonists is a common event in Hepatoblastomas, Clin. Cancer Res. 11 (2005) 4295–4304, https://doi.org/10.1158/1078-0432.CCR-04-1162.
[104] R. John, Y. Atri, V. Chand, N. Jaiswal, K. Raj, A. Nag, Cell cycle-dependent regulation of cytoglobin by Skp2, FEBS Lett. 591 (2017) 3507–3522, https://doi.org/10.1002/1873-3468.12864.
[105] R. John, V. Chand, S. Chakraborty, N. Jaiswal, A. Nag, DNA damage induced activation of Cygb stabilizes p53 and mediates G1 arrest, DNA Repair (Amst) 24 (2014) 107–112, https://doi.org/10.1016/j.dnarep.2014.09.003.
[106] D.V. Hoang, L.T.T. Thuy, H. Hai, V.N. Hieu, K. Kimura, D. Oikawa, Y. Ikura, N. Q. Dat, T.H. Hoang, M. Sato-Matsubara, M.P. Dong, N.V. Hanh, S. Uchida-Kobayashi, F. Tokunaga, S. Kubo, N. Ohtani, K. Yoshizato, N. Kawada, Cytoglobin attenuates pancreatic cancer growth via scavenging reactive oxygen species, Oncogenesis. 11 (2022) 23, https://doi.org/10.1038/s41389-022-00389-4.