Journal of Pathology
J Pathol 2017; **242:** 206–220
Published online 26 April 2017 in Wiley Online Library (wileyonlinelibrary.com) DOI: 10.1002/path.4894

ORIGINAL PAPER

Comparative transcriptome analysis of isogenic cell line models and primary cancers links capicua (CIC) loss to activation of the MAPK signalling cascade

Veronique G LeBlanc[1,2], Marlo Firme[1], Jungeun Song[1], Susanna Y Chan[1], Min Hye Lee[1], Stephen Yip[3], Suganthi Chittaranjan[1] and Marco A Marra[1,4]*

[1] *Canada's Michael Smith Genome Sciences Centre, BC Cancer Agency, Vancouver, BC, Canada*
[2] *Genome Science and Technology Program, University of British Columbia, Vancouver, BC, Canada*
[3] *Department of Pathology and Laboratory Medicine, University of British Columbia, BC, Canada*
[4] *Department of Medical Genetics, University of British Columbia, Vancouver, BC, Canada*

**Correspondence to: MA Marra, Genome Sciences Centre, BC Cancer Agency, 675 West 10th Avenue 7th Floor, Vancouver, BC, Canada V5Z 1L3. E-mail: mmarra@bcgsc.ca*

Abstract

CIC **encodes a transcriptional repressor, capicua (CIC), whose disrupted activity appears to be involved in several cancer types, including type I low-grade gliomas (LGGs) and stomach adenocarcinomas (STADs). To explore human CIC's transcriptional network in an isogenic background, we developed novel isogenic *CIC* knockout cell lines as model systems, and used these in transcriptome analyses to study the consequences of CIC loss. We also compared our results with analyses of transcriptome data from TCGA for type I LGGs and STADs. We identified 39 candidate targets of CIC transcriptional regulation, and confirmed seven of these as direct targets. We showed that, although many CIC targets appear to be context-specific, the effects of CIC loss converge on the dysregulation of similar biological processes in different cancer types. For example, we found that CIC deficiency was associated with disruptions in the expression of genes involved in cell–cell adhesion, and in the development of several cell and tissue types. We also showed that loss of CIC leads to overexpression of downstream members of the mitogen-activated protein kinase (MAPK) signalling cascade, indicating that CIC deficiency may present a novel mechanism for activation of this oncogenic pathway.**

Keywords: capicua; glioma; stomach adenocarcinoma; MAPK signalling

Received 16 February 2017; Revised 27 February 2017; Accepted 9 March 2017

No conflicts of interest were declared.

Introduction

Low-grade gliomas (LGGs) can be separated into three major molecular subtypes that provide superior prognostic information compared to traditional histological classification: type I (*IDH1/2* mutated and 1p/19q co-deleted), type II (*IDH1/2* mutated), and type III (*IDH1/2* wild type) [1–4]. Type I LGGs, which are strongly associated with oligodendrogliomas, are of particular interest because they are associated with better survival, slow growth, and increased chemosensitivity [1]. Hemizygous mutations in the capicua (*CIC*) gene, located on chromosome 19q13.2, are found in ~50–70% of type I LGGs, but are absent from other glioma subtypes [5–8]. Recent studies have indicated that *CIC* mutations are associated with poorer outcome for type I LGGs [9,10]. Multiple distinct *CIC* mutations have also been found within different regions of single lesions [1], indicating that multiple, independently arising *CIC* mutations may contribute to the progression of a single tumour. Together, these observations are compatible with the notion that *CIC* mutations contribute to oncogenic progression in type I LGGs.

CIC was originally identified in *Drosophila melanogaster* as a tissue-specific transcriptional repressor involved in developmental regulation [11–13]. CIC homologues found across metazoans share at least two highly conserved domains: a high mobility group (HMG) box domain involved in DNA binding, and a C-terminal domain (C1) that appears to be necessary for repression in certain contexts in *Drosophila* [14–17]. CIC is a transducer of receptor tyrosine kinase (RTK) signalling that functions through default repression; upon RTK activation, CIC is directly phosphorylated by extracellular signal-regulated kinase (ERK) [11,18], leading to inhibition of CIC activity and de-repression of its target genes. In humans, CIC's most well-characterized target genes are those encoding

the oncogenic transcription factors ETV1, ETV4, and ETV5 [19–21], which have been implicated in several cancer types [22–24].

In this study, we used integrative bioinformatics approaches and novel isogenic cell line models to explore human CIC's transcriptional network. We identified novel candidate targets of CIC regulation, and confirmed some of these as direct targets. We showed that, while CIC appears to have some context-specific activity, CIC deficiency is associated with disruption of similar pathways and processes in biologically distinct contexts, including disruption of cell adhesion-related processes and aberrant overexpression of the mitogen-activated protein kinase (MAPK) signalling cascade.

Materials and methods

Cell culture, cell lysate preparations, and western blot analysis

HEK293a, HOG, and immortalized normal human astrocytes (NHA) cell lines were cultured in Dulbecco's modified Eagle's medium supplemented with 10% (v/v) heat-inactivated fetal bovine serum (Life Technologies, Ottawa, Ontario, Canada). Cell culture was performed in a humidified, 37 °C, 5% CO_2 incubator. Cell lysate preparations and western blot analyses were performed according to standard protocols, which are described in detail in supplementary material, Supplementary materials and methods. Antibody and primer information can be found in supplementary material, Table S1.

Microarray expression profiling

The following biological replicates were analysed: three HEK-derived *CIC* wild type (CIC^{WT}) lines (HEK, F12, and B7) and three HEK-derived *CIC* knockout (CIC^{KO}) lines (D10, A9, and D1); and three separate passages each of the parental CIC^{WT} (HOG) line and of the HOG-derived CIC^{KO} (F11) line. RNA extraction was performed with the RNeasy Plus Mini Kit (Qiagen, Montreal, Quebec, Canada), according to the manufacturer's recommendations. Microarray expression profiling was performed with the GeneChip Human Gene 2.0 ST array (Affymetrix, Santa Clara, CA, USA) at the Centre for Applied Genomics, The Hospital for Sick Children, Toronto, Canada. Robust multichip average (RMA) normalization was performed with the R/Bioconductor package oligo [25] (version 1.34.2), with gene-level summarization of core probeset data. Annotation was performed with the R/Bioconductor package hugene20sttranscriptcluster.db (version 8.5.0), and only transcript clusters that mapped to single genes were retained for further analyses. Multiple transcript clusters that mapped to identical genes were aggregated by the use of median expression values. To identify candidate target genes, fold-change (FC) differences in gene expression were calculated for each gene between each individual CIC^{KO}/CIC^{WT} pair. Genes with an FC value of >1.5 in at least four (HEK) or six (HOG) comparisons were considered to be differentially expressed (DE) [26]. The data are accessible through the Gene Expression Omnibus (dataset GSE80359).

TCGA expression analyses

RNA-sequencing results were obtained from the TCGA data portal (https://tcga-data.nci.nih.gov/tcga/; see supplementary material, Table S2, for sample information). Motivated by our observation that a proportion of CIC^{WT} samples in type I LGGs showed relatively low *CIC* mRNA expression (supplementary material, Figure S1), and given the possibility that alterations other than sequence variants could affect *CIC* expression [27], we analysed data from CIC^{WT} samples with *CIC* expression greater than the first quartile, giving a total of 68 CIC^{WT} samples and 39 samples with truncating *CIC* mutations. For stomach adenocarcinoma (STAD), samples with a *CIC* copy number loss (CIC^{loss}, $n = 48$) were compared to samples with intact *CIC* ($n = 155$). Samples with a *CIC* mutation were excluded. The R/Bioconductor package DESeq2 [28] (version 1.10.0) was used to conduct differential expression analyses.

Results

Transcriptome analysis of CIC^{KO} cell line models identifies known and novel candidate targets of CIC transcriptional regulation

In an effort to minimize the confounding effects of the multiple mutations found in cancer genomes and their impacts on the transcriptome, we generated isogenic CIC^{KO} cell lines by using a zinc finger nuclease [29] and the CRISPR/Cas9 [30,31] technology in HEK293a (HEK) and glioma-derived HOG cells [32] (supplementary material, Figure S2A). Both approaches were designed to produce insertions or deletions within exon 2, which is shared between the short (CIC-S) and long (CIC-L) CIC isoforms [33] (supplementary material, Figure S2B). Three HEK-derived monoclonal cell lines and one HOG-derived monoclonal cell line with undetectable CIC expression were obtained (Figure 1A, B; supplementary material, Figure S2C). We functionally validated the CIC^{KO} lines by measuring the expression of the known direct CIC targets *ETV1*, *ETV4*, and *ETV5* [19–21]. The HEK-derived CIC^{KO} lines had significant increases in *ETV1/4/5* expression relative to the CIC^{WT} controls, and the HOG-derived CIC^{KO} line showed similar trends, particularly for *ETV4* (Figure 1C). Together, the lack of detectable CIC protein expression and the increased expression of known CIC targets indicated that our CIC^{KO} lines had lost CIC's transcriptionally repressive function.

We next performed microarray gene expression analyses on our cell line models to identify genes whose expression was affected by CIC loss (supplementary

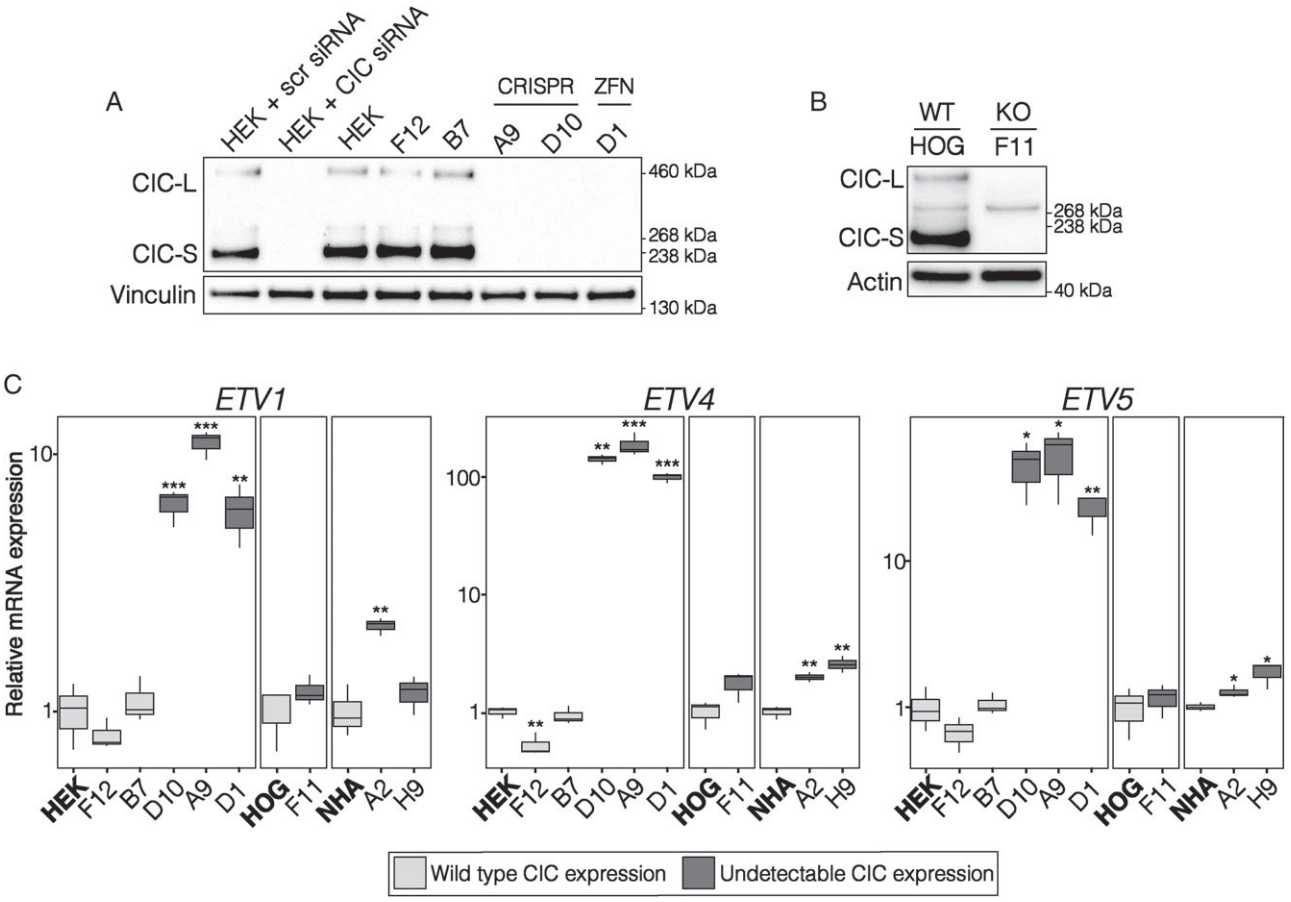

Figure 1. Novel *CIC*^{KO} cell line models lack functional CIC. (A) Representative western blot of HEK-derived *CIC*^{WT} (HEK, F12, and B7) and *CIC*^{KO} (A9, D10, and D1) cell lines profiled by the use of microarrays. A9 and D10 were obtained using the CRISPR/Cas9 technology, and D1 was obtained using a zinc finger nuclease. HEK + siRNA: HEK293a cells treated with a 'scrambled' non-targeting control (scr) or *CIC*-specific siRNA to confirm CIC antibody specificity. Vinculin was used as a loading control. (B) Representative western blot of the HOG cell line and its *CIC*^{KO} derivative (F11). Actin was used as a loading control. (C) Tukey boxplots showing relative *ETV1/4/5* mRNA expression, as measured by reverse transcription (RT)-qPCR, in the indicated cell lines compared to their respective parental cell line (in bold). Data were obtained from three independent experiments. $^{*}p < 0.05$, $^{**}p < 0.01$ and $^{***}p < 0.001$ relative to the parental cell line (two-sided Student's *t*-test).

material, Tables S3 and S4). Interestingly, although CIC has been observed to function as a transcriptional repressor, our list of candidate targets obtained from the HEK dataset was approximately equally distributed between genes that showed overexpression (427 of 929 genes, 46%) and underexpression (502 of 929 genes, 54%) in *CIC*^{KO} lines (supplementary material, Table S3). Of note, the HOG dataset showed a more skewed distribution, with 411 of 611 genes (67%) showing higher expression in the *CIC*^{KO} line (supplementary material, Table S4). While the HEK-derived *CIC*^{KO} lines showed increased expression of the known CIC targets *ETV1*, *ETV4*, and *ETV5*, only *ETV4* passed the threshold for increased expression in the HOG-derived *CIC*^{KO} lines, although *ETV5* showed a similar trend. Together, these results indicate that CIC may have some context-dependent targets and/or activity. Interestingly, previous studies that have compared transcriptome profiles of type I LGGs have identified either a majority of downregulated genes (66%) [10] or exclusively upregulated genes [34] in *CIC* mutant samples (Table 1) [35,36].

To gain insights into the biological role of CIC loss and its associated dysregulated gene expression patterns, we performed functional enrichment analyses. Biological processes significantly enriched for DE genes, classified into clusters of terms defined by similar gene sets, were dominated by those related to central nervous system (CNS) development and regulation (9/40, supplementary material, Table S5A). While this reflects CIC's role in nervous system development [16,17], several clusters were also related to the development of other organs and systems, including the kidney, mammary gland, female reproductive system, and bone and vasculature. Given that CIC has been implicated in the development of several organ sites in *Drosophila* [13,37–39] and mice [17,40,41], these results indicate that CIC may also play a more widespread and extensive role in human development than currently appreciated. Terms related to cell migration, chemotaxis, extracellular matrix organization, and cell adhesion may provide further insights into the mechanism by which CIC loss contributes to increased metastatic potential in lung cancer cells [35]. Notably, several gene families had multiple members represented in these terms, such

Table 1. Overlap with previously identified candidate targets of CIC transcriptional regulation

	Glioma [10,34]	Lung cancer cell lines [35]	Prostate cancer cell lines [36]
HEK	***CCND1***, ***DUSP4***, ***DUSP6***, ***ETV1***, ***ETV4***, ***ETV5***, *FLRT3*, ***GPR3***, *HPCAL4*, *OLIG2*, *PLPPR5*, *PPP1R14C*, ***PPP2RC2***, ***ROBO2***, ***SHC3***, ***SLC35F1***, *SOX11*, ***SPRED1***, ***SPRED2***, ***SPRY4***, ***TMOD1***, ***TRAPPC9***	*CKMT1A*, ***ETV4***, ***ETV5***, *KRT19*, *MYO10*, ***PRAME***, ***PTPN9***, ***SPRED1***, *SERPINB9*, *SOX4*, ***ZNF486***	***CRABP1***, ***CTGF***, *IFI44L*, *LINC01116*, *MIR570*, ***PPP2R2C***, ***TPD52L1***, *VCAN*, *VTRNA1-2*, *ZNF702P*
HOG	*C3orf80*, ***DUSP6***, ***ETV4***, ***GPR3***, *ICA1*, *LRRC4C*, ***NRG1***, ***RAB31***, ***ROBO2***, ***SPRY4***, *STAMBPL1*, *STC2*	*COTL1*, ***CRABP2***, *DZIP3*, ***ETV4***, ***FAS***, ***NRG1***, *NUAK2*, ***NUDT7***, *PPARG*, *PRTFDC1*, *SULT2B1*, *TCEAL1*, *TBL1X*	***ADAMTS1***, *BHLHE41*, *CCDC15*, ***COL8A1***, ***CRABP2***, ***CTGF***, *EPGN*, *GMPR*, *IL22RA1*, ***MOXD1***, *NPY1R*, *PKIB*, ***RAB31***, *SNAI2*, *TBC1D1*, ***TMEM171***, ***TPD52L1***, *UCP2*
Type I LGG	*ANKRD55*, *BAALC*, *BCL2*, ***BACH2***, *C2orf27A*, *C3orf31*, *C6orf118*, *C8orf56*, *CADPS*, *CAMK2N1*, ***CCND1***, *CD82*, ***CNTNAP4***, ***CREB3L1***, *DIAPH2*, ***DCLK1***, *DLL3*, ***DUSP4***, ***DUSP6***, *ELFN1*, *EPN2*, *ESRRG*, *ETS1*, ***ETV1***, ***ETV4***, ***ETV5***, *FBFBP3*, *FGFR1*, *FOXP4*, *GCNT2*, *GFRA1*, ***GLDC***, *GLT25D2*, ***GPR3***, ***IPO8***, *KCNIP1*, *KCNK3*, *KIAA1598*, *LASP1*, *LBH*, ***LMO1***, ***LOC92659***, *LPPR5*, *MGC12982*, *NCAN*, *NLGN3*, *NPPA*, ***NRG1***, *NUDT9P1*, *PEX5L*, ***PDE4B***, *PDGFRA*, ***RAB31***, *RASGRF1*, *RNF216L*, ***SCARA5***, *SEMA4D*, *SIX1*, *SCEL*, ***SHC3***, *SPOCK3*, *SLC29A1*, ***SLC35F1***, ***SPRED1***, ***SPRED2***, ***SPRY4***, *SPSB4*, *TACC2*, *TMC3*, *TMEM158*, ***TMOD1***, *TRAF4*, ***TRAPPC9***, *TRIB2*, *TTLL7*, *UHRF1*, *VSIG10*, *WSCD1*, *ZBTB8B*, *ZSWIM4*	*CNP*, ***ETV4***, ***ETV5***, *HAS3*, *HEXIM2*, *ID4*, ***IPO8***, *LPGAT1*, ***NRG1***, *NRTN*, ***NUDT7***, *PAIP2B*, ***PDE4B***, ***PTPN9***, *SKAP2*, ***SPRED1***, *TM4SF18*, *YWHAQ*	***CRABP1***, ***CREB3L1***, *GPR4*, *LRIG1*, *MARCH9*, ***MOXD1***, *MT1G*, *MT1L*, *PLA2G1B*, *PPL*, *PRPH*, ***RAB31***, *ROBO4*, ***SCARA5***, ***TMEM171***, ***TPD52L1***
High-confidence candidate targets	***CCND1***, ***DUSP4***, ***DUSP6***, ***ETV1***, ***ETV4***, ***ETV5***, ***PLPPR5***, ***RAB31***, ***SHC3***, ***SPRED1***, ***SPRED2***, ***SPRY4***, ***TRAPPC9***	***NUDT7***, ***PTPN9***, ***SPRED1***	***CRABP1***, ***RAB31***, ***TPD52L1***
STAD	*ADAMTS2*, *ALK*, ***BACH2***, *BAALC*, *BCL2*, *BMPER*, ***CNTNAP4***, ***DCLK1***, *DPP6*, ***ETV4***, *FAM65B*, *FKBP5*, ***GLDC***, *GPR17*, *ISM1*, *KCND2*, *KLF9*, ***LMO1***, ***LOC92659***, *LRRC7*, *NRXN2*, *NXPH3*, ***PDE4B***, ***SCARA5***, *SFRP1*, *SHROOM2*, *SNCAIP*, *TMEM132C*, ***TMOD1***	*ATP2B4*, *C11orf86*, *CREB3L3*, *CRISPLD2*, *DPYSL3*, ***ETV4***, ***FAS***, *HEYL*, ***PDE4B***, ***PRAME***, *PRX*, *TGFB3*, *S100A9*, *ZCCHC24*, *ZNF217*, ***ZNF486***, *ZNF772*	*ADAM12*, ***ADAMTS1***, *AK5*, *ARHGDIB*, *C1R*, *C1S*, *COL6A3*, ***COL8A1***, ***CRABP1***, *FAM107A*, *GAS6*, *GHRL*, *GLI3*, *HCLS1*, *HIST1H2BH*, *HLF*, *LCP1*, ***MOXD1***, *OPRL1*, *PLEKHO1*, *PRPH*, *RUNDC3B*, *SCARA5*, *SERPINB2*, *TGFB1*

LGG, low-grade glioma; STAD, stomach adenocarcinoma.
The genes identified in this study (rows) as candidate targets of CIC transcriptional regulation overlap with previously identified candidate targets (columns) in biologically distinct contexts. Genes in bold are found in more than one condition (row or column).

as protocadherin (*PCDH*) genes (which were universally underexpressed in CIC^{KO} lines), and semaphorin, collagen, and annexin genes. Hallmark gene sets and oncogenic signatures significantly enriched for genes overexpressed in CIC^{KO} lines included gene sets whose expression was found to increase upon activation of epidermal growth factor receptor (EGFR), ERBB2, RAF, KRAS, MEK, or mammalian target of rapamycin (MTOR) (supplementary material, Table S5B), implicating CIC in the control of these signalling pathways. Similarly, signatures significantly enriched for genes underexpressed in CIC^{KO} lines included gene sets whose expression was found to decrease upon activation of KRAS, MEK, or MTOR, and upon knockdown of *RB*, *E2F1*, or *P53* (supplementary material, Table S5C).

Transcriptome analysis of type I LGGs identifies high-confidence candidate targets of CIC transcriptional regulation

To explore the consequences of CIC deficiency in a primary tumour context, we obtained RNA-sequencing data for type I LGGs from TCGA [42]. Hemizygous *CIC* mutations found in type I LGGs show an interesting pattern, whereby ~50% are truncating mutations distributed throughout the gene, and the remainder are missense mutations that cluster within the conserved HMG domain (Figure 2A) [42]. To assess whether this distribution could be correlated with different patterns of transcriptional dysregulation, we analysed the expression of known CIC targets within tumour samples with missense (CIC^{mis}) or truncating (CIC^{trunc}) mutations. As expected, the expression of *ETV1/4/5* was significantly higher in *CIC* mutant samples than in CIC^{WT} samples, regardless of mutation type (Figure 2B). However, *ETV4* also showed significantly higher expression in CIC^{trunc} than in CIC^{mis} samples, and a similar trend was observed for *ETV5*, indicating that CIC missense mutants may retain some repressive activity. To explore this possibility, we transfected CIC^{KO} cells with FLAG-tagged *CIC* constructs together with a luciferase reporter designed to drive expression through the *ETV5* promoter sequence (supplementary material, Figure S3). Luciferase activity following reintroduction of *CIC* constructs with missense mutations was reduced similarly to luciferase activity following reintroduction

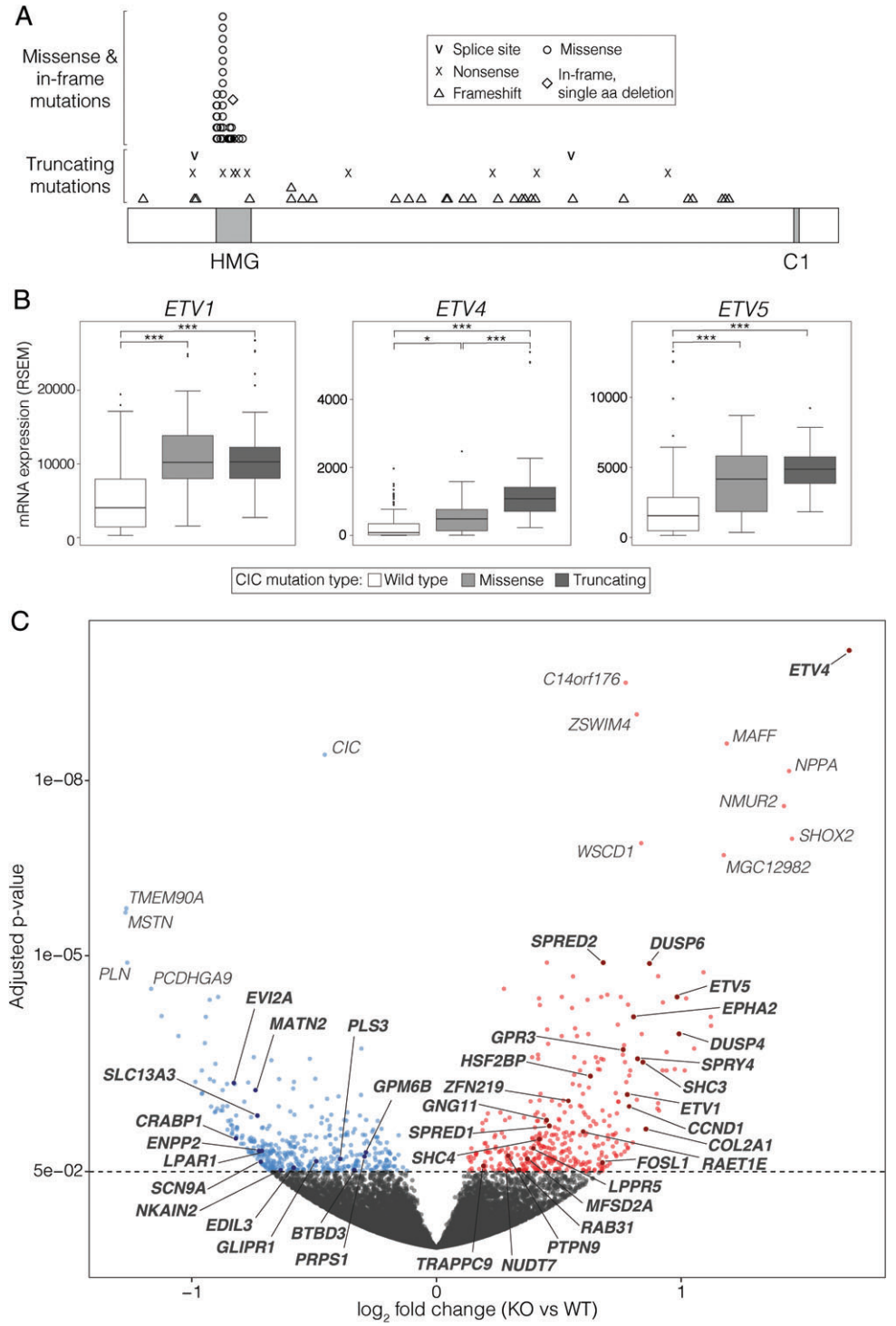

Figure 2. Transcriptome profiling identifies known and novel candidate targets of CIC transcriptional regulation. (A) Distribution of *CIC* mutations found in 78 type I LGG samples with *CIC* mutations from TCGA (supplementary material, Table S2). (B) Tukey boxplots showing gene expression for *ETV1*, *ETV4* and *ETV5* in type I LGGs from TCGA for samples with wild-type *CIC* expression ($n = 91$), missense *CIC* mutations ($n = 38$), and truncating *CIC* mutations ($n = 39$). $^{*}p < 0.05$ and $^{***}p < 0.001$ (two-sided Student's *t*-test) (C) Volcano plot of gene expression in type I LGGs with truncating *CIC* mutations ($n = 39$) compared to those with wild-type *CIC* and high *CIC* expression ($n = 68$). High-confidence candidate target genes (see Results) are labelled in bold (Table 2).

of CIC^{WT}, confirming that the mutant constructs retain some repressive activity. Conversely, reintroduction of a truncated form of *CIC* did not affect luciferase activity, consistent with complete loss of CIC's repressive activity.

We therefore studied CIC's transcriptional network within the context of LGGs, comparing CIC^{trunc} ($n = 39$) with CIC^{WT} ($n = 68$) samples. A differential expression analysis identified 799 DE genes (FDR of < 5%; Figure 2C; supplementary material, Table S6). Although a similar analysis was performed previously [34], ours

considered 84 additional samples and updated mutational annotations, in which the status of eight samples changed from *CIC* mutant to *CIC*^{WT}. Furthermore, whereas this earlier study exclusively reported genes showing increased expression in *CIC* mutant samples, our DE genes were approximately equally distributed between genes showing overexpression and underexpression in *CIC*^{trunc} samples [380/799 (48%) and 419/799 (52%), respectively], which is consistent with the results obtained in our cell line models.

To identify genes whose differential expression was consistently associated with CIC loss, we analysed the overlap between DE genes obtained from our *CIC*^{KO} lines and from type I LGGs (Table 2). Of the 58 genes that showed differential expression in primary tumour samples and in at least one cell line model, 39 (67%) had consistent directional changes (shaded in Table 2; Figure 2C). These 39 genes included the known CIC target genes *ETV1*, *ETV4*, and *ETV5*, along with 14 other genes previously reported to be candidate CIC targets (Table 1), and were considered to be high-confidence candidate targets of CIC transcriptional regulation. It is of note that *ETV4*, *DUSP6*, *SPRY4* and *GPR3* showed increased expression in all three contexts. Importantly, the 19 genes that did not show consistent directional changes in expression may still represent direct or indirect targets of CIC, as CIC's transcriptional regulation activity may be, at least in part, context-dependent. CIC's possible context dependency is further supported by the absence of an increase in the known targets *ETV1* and *ETV5* seen only in the HOG *CIC*^{KO} lines.

High-confidence candidate targets show evidence of CIC regulation in isogenic cell line models

To confirm the expression changes described above, a subset of the high-confidence candidate targets were further validated at the mRNA and protein levels in the HEK-derived and HOG-derived *CIC*^{KO} lines, along with additional *CIC*^{KO} lines derived from a normal human astrocyte (NHA) line stably expressing wild-type *IDH1* [43] (Figure 1C). mRNA levels for *GPR3*, *SPRED1*, *SHC3*, and *SHC4* were significantly increased in HEK-derived *CIC*^{KO} lines, and *DUSP4* and *DUSP6* showed similar trends (Figure 3A). *GPR3*, *SPRED1*, *SHC4*, *DUSP4*, and *DUSP6* also had significantly increased expression in the HOG-derived *CIC*^{KO} line, and all genes tested showed similar trends in the NHA-derived *CIC*^{KO} lines, reaching significance for *SPRED1* and *DUSP4*. Gene expression results were also confirmed by western blots, with SPRY4, LRP8, DUSP6, and PTPN9 showing significantly increased expression in HEK-derived *CIC*^{KO} lines (Figure 3B and supplementary materials, Figure S4), and ETV4, SPRY4 and DUSP6 showing increased expression in the HOG-derived *CIC*^{KO} line (Figure 3C).

To confirm that the increased protein expression of candidate targets is attributable to loss of CIC, we reintroduced *CIC* into one of the *CIC*^{KO} lines. ETV4, SPRY4, and DUSP6 showed reduced expression upon reintroduction of *CIC*, but not upon introduction of an empty FLAG construct (Figure 3D), indicating that reintroduction of *CIC* is sufficient to suppress their expression. Interestingly, LRP8 expression remained similar upon reintroduction of *CIC*; given that CIC can function with a co-repressor in *Drosophila* [11,44], it is conceivable that a similar interaction occurs in humans, possibly also in a context-dependent manner, and that this may be needed for effective repression of some of CIC's target genes. These results indicate that loss of CIC has potentially oncogenic functional consequences beyond transcriptional expression changes.

Promoter regions associated with candidate target genes show evidence of CIC binding

To gauge whether the candidate CIC targets identified by our analyses were likely to be direct targets, we analysed their promoter regions [defined as 1500 bp upstream and 500 bp downstream of the transcription start site (TSS)] [45] to identify putative CIC binding sites. To do this, we made use of a previously defined CIC octameric consensus binding site (TG/CAATGG/AG/A; Figure 4A) [46]. We performed our analyses allowing for one mismatch at position 2, 7, or 8 (i.e. the positions where sequence frequency is <100%). Genes identified as being DE in the HOG-derived *CIC*^{KO} lines (611 genes) or in *CIC*^{trunc} type I LGGs (799 genes) were found to harbour significantly more of these putative binding sites in their promoters than genes showing no differences in expression (Fisher's exact test: $p = 0.043$ and $p = 5.44 \times 10^{-5}$, respectively). The 929 genes identified as being DE in the HEK-derived *CIC*^{KO} lines showed a similar trend ($p = 0.090$). Notably, high-confidence candidate target genes were also associated with promoter regions that were significantly enriched for these putative binding sites ($p = 0.036$), indicating that they are likely to be enriched for direct targets. This notion is further supported by the presence within this list of CIC's known direct targets (*ETV1*, *ETV4*, and *ETV5*), whose promoters contain seven to 15 of these putative binding sites (Table 3).

To confirm CIC binding in the promoter region of a subset of the high-confidence candidate target genes, we performed targeted chromatin immunoprecipitation (ChIP) followed by quantitative polymerase chain reaction (qPCR) analysis. Putative CIC binding sites in the promoter regions of *ETV4*, *GPR3*, *DUSP4*, *DUSP6*, *SHC3*, *SHC4*, *SPRY4*, and *SPRED1* showed significant enrichment (~2.5–80-fold differences, $p < 0.05$) as compared with a negative control region (NCR) (NCR1; Figure 4B). Interestingly, three of the four sites tested in the *ETV4* promoter region showed significant enrichment (~40–60-fold differences), including one site in the promoter region of the shorter *ETV4* isoform (site D, uc002idv.5; supplementary material, Figure S5B). However, a second site in this same region (site C) did not show any enrichment, despite containing the same consensus sequence (TGAATGAA) as sites A and B. Of

Table 2. High-confidence candidate targets of CIC transcriptional regulation

Transcript cluster ID	Entrez ID	Gene	Description	Chromosome	HEK microarray			HOG microarray			TCGA type I LGG		
					Log_2 fold change*	Up frequency†	Down frequency†	Log_2 fold change*	Up frequency†	Down frequency†	Log_2 fold-change	*P* value	BH-adjusted *P* value
16845410	2118	*ETV4*	ETS variant 4	17	3.87	9	0	0.89	7	0	1.69	3.30E-15	5.94E-11
16898403	200734	*SPRED2*	Sprouty-related EVH1 domain-containing 2	2	0.65	6	0	0.22	0	0	0.68	9.94E-09	1.28E-05
16768297	1848	*DUSP6*	Dual-specificity phosphatase 6	12	0.95	8	0	1.08	8	0	0.87	1.21E-08	1.36E-05
16962380	2119	*ETV5*	ETS variant 5	3	2.42	9	0	0.53	3	0	0.99	7.62E-08	5.08E-05
17012087	222553	*SLC35F1*	Solute carrier family 35 member F1	6	−0.69	3	5	0.16	0	0	0.67	6.55E-08	5.08E-05
16682098	1969	*EPHA2*	EPH receptor A2	1	0.87	6	0	0.02	0	0	0.81	2.29E-07	1.11E-04
17075973	1846	*DUSP4*	Dual-specificity phosphatase 4	8	0.81	7	0	−0.02	0	0	0.99	5.33E-07	2.18E-04
16661429	2827	*GPR3*	G-protein-coupled receptor 3	1	2.33	9	0	1.21	8	0	0.76	1.24E-06	4.05E-04
17001063	81848	*SPRY4*	Sprouty RTK signalling antagonist 4	5	2.04	9	0	0.77	6	0	0.82	2.14E-06	5.82E-04
17095566, 17095582	53358	*SHC3*	SHC adaptor protein 3	9	0.78	5	0	−0.31	0	0	0.85	2.62E-06	6.65E-04
16843167	2123	*EVI2A*	Ecotropic viral integration site 2A	17	−0.06	0	0	−2.36	0	9	−0.83	7.89E-06	1.51E-03
17071218	4147	*MATN2*	Matrilin 2	8	−0.20	1	4	0.79	9	0	−0.74	1.14E-05	2.00E-03
17055354	2115	*ETV1*	ETS variant 1	7	2.32	9	0	−0.03	0	0	0.78	1.45E-05	2.39E-03
16790202	51222	*ZNF219*	Zinc finger protein 219	14	0.54	4	0	0.18	1	0	0.54	1.99E-05	3.06E-03
16728261	595	*CCND1*	Cyclin D1	11	0.51	4	0	−0.37	0	0	0.79	2.67E-05	3.80E-03
17101923	5475	*PPEF1*	Protein phosphatase with EF-hand domain 1	X	0.80	6	0	0.05	0	0	−0.88	3.42E-05	4.45E-03
16919879	64849	*SLC13A3*	Solute carrier family 13 member 3	20	−0.59	0	5	0.17	1	0	−0.73	4.70E-05	5.44E-03
17048450	2791	*GNG11*	G-protein subunit gamma 11	7	0.35	4	0	−0.01	0	0	0.45	6.12E-05	6.55E-03
16799231	161742	*SPRED1*	Sprouty-related EVH1 domain-containing 1	15	0.92	9	0	0.37	0	0	0.46	8.74E-05	8.23E-03
16763783	1280	*COL2A1*	Collagen type II alpha 1 chain	12	0.52	4	1	0.04	0	0	0.86	1.04E-04	9.31E-03
17024669	135250	*RAET1E*	Retinoic acid early transcript 1E	6	0.39	4	0	−0.23	0	0	0.60	1.21E-04	1.03E-02
17068014	23259	*DDHD2*	DDHD domain-containing 2	8	0.10	0	0	0.88	9	0	−0.24	1.85E-04	1.28E-02
16803492	1381	*CRABP1*	Cellular retinoic acid-binding protein 1	15	−0.46	0	4	0.24	0	0	−0.82	2.01E-04	1.34E-02
16808918	399694	*SHC4*	SHC adaptor protein 4	15	1.11	9	0	0.18	0	0	0.42	2.15E-04	1.39E-02
17085760	256691	*MAMDC2*	MAM domain-containing 2	9	0.70	5	0	−0.08	0	0	−0.63	2.24E-04	1.44E-02
16762978	196394	*AMN1*	Antagonist of mitotic exit network 1 homologue	12	−0.06	0	0	0.90	9	0	−0.23	2.74E-04	1.64E-02
17080516	5168	*ENPP2*	Ectonucleotide pyrophosphatase/phosphodiesterase 2	8	−0.89	1	5	−0.10	0	0	−0.73	4.57E-04	2.19E-02
17097132	1902	*LPAR1*	Lysophosphatidic acid receptor 1	9	0.03	0	0	−0.88	0	9	−0.71	4.62E-04	2.21E-02
17023658	26002	*MOXD1*	Monooxygenase DBH-like 1	6	−0.21	0	2	0.92	8	0	−0.76	4.92E-04	2.27E-02
17109170	2824	*GPM6B*	Glycoprotein M6B	X	−0.61	0	4	−0.15	0	0	−0.29	5.28E-04	2.34E-02
16920299	3755	*KCNG1*	Potassium voltage-gated channel modifier subfamily G member 1	20	−0.01	0	0	0.66	6	0	−0.73	6.22E-04	2.54E-02
16811816	5780	*PTPN9*	Protein tyrosine phosphatase, non-receptor type 9	15	0.59	4	0	0.27	1	0	0.29	6.84E-04	2.66E-02
17106031	5631	*PRPS1*	Phosphoribosyl pyrophosphate synthetase 1	X	0.21	0	0	−0.67	0	6	−0.29	7.17E-04	2.76E-02
17106357	5358	*PLS3*	Plastin 3	X	−0.10	1	1	−0.67	0	8	−0.39	8.66E-04	3.03E-02
16662945	84879	*MFSD2A*	Major facilitator superfamily domain-containing 2A	1	0.86	7	0	0.00	0	0	0.37	8.70E-04	3.03E-02
17108799	415	*ARSE*	Arylsulfatase E (chondrodysplasia punctata 1)	X	−0.83	1	4	0.14	0	0	0.48	9.11E-04	3.13E-02
16754373	11010	*GLIPR1*	GLI pathogenesis-related 1	12	−0.09	0	1	−0.67	0	6	−0.49	9.91E-04	3.32E-02

Table 2. Continued

Transcript cluster ID	Entrez ID	Gene	Description	Chromosome	HEK microarray			HOG microarray			TCGA type I LGG		
					Log_2 fold change*	Up frequency†	Down frequency†	Log_2 fold change*	Up frequency†	Down frequency†	Log_2 fold-change	*P* value	BH-adjusted *P* value
16904667	6335	*SCN9A*	Sodium voltage-gated channel alpha subunit 9	2	−0.64	0	5	−0.06	0	0	−0.72	1.05E-03	3.39E-02
17095503	440173	*LOC440173*	Uncharacterized LOC440173	9	−0.55	0	5	0.07	0	0	0.60	1.10E-03	3.49E-02
16740630	8061	*FOSL1*	FOS-like 1, AP-1 transcription factor subunit	11	0.90	6	0	−0.46	0	3	0.68	1.11E-03	3.50E-02
17087329	7111	*TMOD1*	Tropomodulin 1	9	−0.51	0	4	0.09	1	0	0.38	1.13E-03	3.56E-02
17075448	8793	*TNFRSF10D*	TNF receptor superfamily member 10d	8	−0.86	0	5	0.40	2	0	0.47	1.17E-03	3.62E-02
17012281	7164	*TPD52L1*	Tumour protein D52-like 1	6	−0.86	0	6	0.67	6	0	0.57	1.21E-03	3.69E-02
17067696	3084	*NRG1*	Neuregulin 1	8	−0.20	0	1	−1.33	0	9	0.60	1.24E-03	3.73E-02
16986065	134285	*TMEM171*	Transmembrane protein 171	5	−0.09	0	0	1.22	8	0	−0.68	1.27E-03	3.79E-02
17050522	26136	*TES*	Testin LIM domain protein	7	−0.13	0	0	−0.83	0	8	0.44	1.37E-03	3.96E-02
17081665	83696	*TRAPPC9*	Trafficking protein particle complex 9	8	0.48	4	0	0.33	1	0	0.19	1.40E-03	4.00E-02
16793692	27133	*KCNH5*	Potassium voltage-gated channel subfamily H member 5	14	−0.11	0	0	0.89	9	0	−0.70	1.49E-03	4.10E-02
16819257	4496	*MT1H*	Metallothionein 1H	16	−0.59	0	6	0.25	0	0	0.69	1.54E-03	4.16E-02
16997816	10085	*EDIL3*	EGF-like repeats and discoidin domains 3	5	−0.53	3	4	0.77	7	0	−0.58	1.61E-03	4.29E-02
16850923	11031	*RAB31*	RAB31, member RAS oncogene family	18	−0.26	0	2	0.90	8	0	0.34	1.84E-03	4.63E-02
16682402	11240	*PADI2*	Peptidyl arginine deiminase 2	1	0.44	4	0	0.90	7	0	−0.60	1.88E-03	4.65E-02
17012245	154215	*NKAIN2*	Na^+/K^+-transporting ATPase-interacting 2	6	−0.44	0	4	0.10	0	0	−0.64	1.96E-03	4.72E-02
16911463	22903	*BTBD3*	BTB domain-containing 3	20	−0.57	0	4	0.00	0	0	−0.34	1.98E-03	4.74E-02
17104416	3476	*IGBP1*	Immunoglobulin (CD79A)-binding protein 1	X	0.17	0	0	−0.79	0	8	0.23	2.00E-03	4.74E-02
16887147	129446	*XIRP2*	Xin actin-binding repeat-containing 2	2	−0.34	0	4	−0.22	0	0	0.65	2.06E-03	4.81E-02
16821162	283927	*NUDT7*	Nudix hydrolase 7	16	−0.15	0	1	1.18	9	0	0.29	2.09E-03	4.85E-02
17061099	10156	*RASA4*	RAS p21 protein activator 4	7	0.23	4	2	−0.15	1	2	0.42	2.19E-03	4.97E-02

BH, benjamini-hochberg; LGG, low-grade glioma; RTK, receptor tyrosine kinase; TNF, tumour necrosis factor.

Shaded genes have consistent directional changes in expression in either HEK or HOG and Type I LGG, and are considered to be high-confidence candidate targets of CIC transcriptional regulation. Genes are ranked by their BH-adjusted *P* values in Type I LGGs.

*Average of all pairwise comparisons.

†Frequency indicates the number of pairwise comparisons with a fold change value of >1.5 (up) or <-1.5 (down).

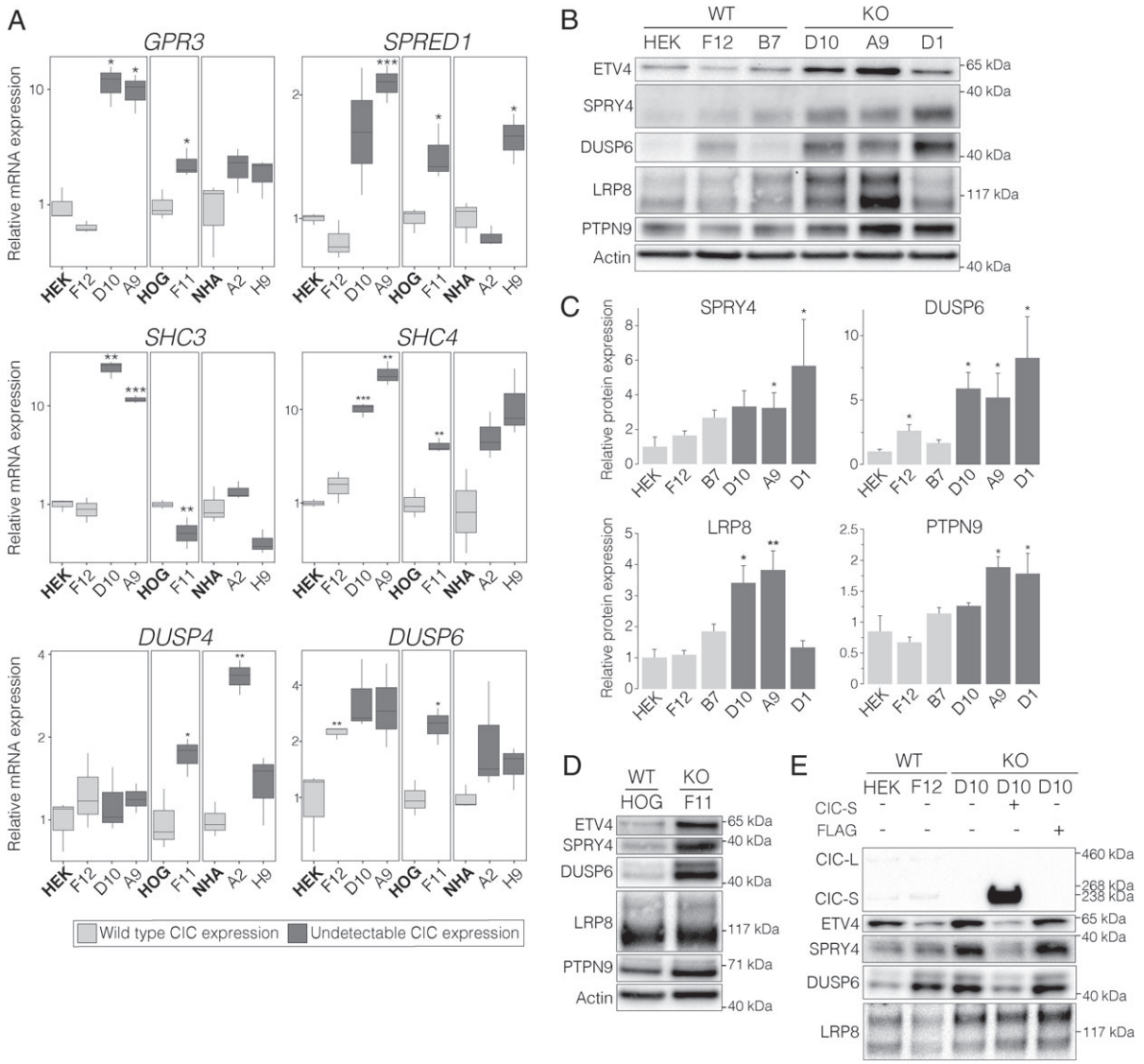

Figure 3. High-confidence candidate targets of CIC regulation show increased transcript and protein expression in *CIC*^{KO} cells. (A) Tukey boxplots showing expression of candidate target genes, as measured by RT-qPCR, in the indicated cell lines compared to their respective parental cell lines (in bold). (B) Representative western blots showing increased expression of candidate CIC target genes in HEK-derived *CIC*^{KO} lines compared to *CIC*^{WT} lines. Actin was used as a loading control, and a representative blot is shown. (C) Quantification of western blots for candidate CIC targets, showing mean relative expression compared to HEK cells. Additional quantifications are shown in supplementary material, Figure S4. All quantifications in (A) and (C) were obtained from three independent experiments. Error bars (C): standard error of the mean. $^{*}p < 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$ (two-sided Student's *t*-test). (D) Representative western blots showing increased expression of candidate CIC target genes in the HOG-derived *CIC*^{KO} cell line compared to the parental cell line. (E) Representative western blots showing decreased expression of ETV4, SPRY4, and DUSP6 in a *CIC*^{KO} cell line following reintroduction of *CIC*. A FLAG construct lacking *CIC* was used as a control.

the other sites that did not show evidence of CIC binding, only half (3/6) had a single-base mismatch to the CIC consensus sequence (*PTPN9* site A, TGAATGAT; *SHC4* site A, TAAATGGA; and *SPRED2* site A, TGAATGTG). However, two sites with a mismatch (*DUSP6* site C and *SHC3* site A, TTAATGAG) did show significant enrichment, suggesting that CIC can still bind in the presence of a mismatch, and may particularly tolerate a T at position 2. Importantly, CIC binding affinity may be further influenced by contextual elements such as the surrounding sequence, distance to the TSS, or cofactor binding; however, further genome-wide studies will be needed to investigate these possibilities.

CIC deficiency in biologically distinct contexts leads to dysregulation of similar pathways

CIC aberrations have recently begun to be associated with additional cancer types, such as sarcomas [19,47], prostate cancer [36], and lung cancer [35]. *CIC* is also significantly mutated in microsatellite instability (MSI) subtype STADs [48], and decreased CIC expression was found to correlate with disease stage in STAD samples, while overexpression of wild-type CIC in a *CIC*^{mis} STAD cell line decreased its invasive potential [35]. To further characterize CIC's transcriptional network within distinct contexts and to investigate whether similar genes were affected by CIC deficiency in different

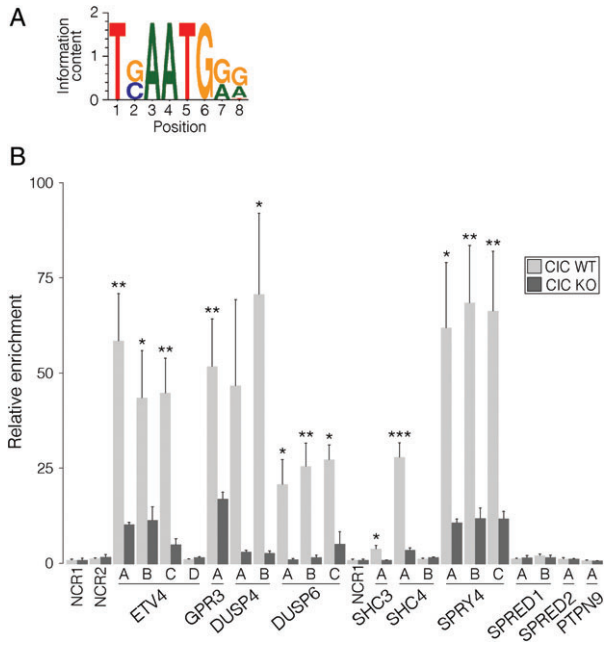

Figure 4. Promoter regions of high-confidence candidate targets of CIC regulation show enrichment of CIC binding. (A) Consensus CIC binding sequence logo [11]. (B) Mean enrichment of putative CIC binding sites relative to NCR1 following ChIP-qPCR for CIC in *CIC*^{WT} (HEK) and *CIC*^{KO} (D10) cell lines. More detailed information can be found in supplementary material, Figure S5. Error bars: standard error of the mean over four (*CIC*^{WT}) or three (*CIC*^{KO}) independent experiments. qPCR analyses for each replicate had to be performed on two plates, and respective NCR1 values are shown. $^{*}p < 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$ (two-sided Student's *t*-test).

cancer types, we identified genes whose differential expression was associated with loss of CIC in STAD [48]. This yielded 1924 DE genes, including *ETV4* (FDR of <5% and FC of >1.5; supplementary material, Tables S7 and S8).

To determine whether similar processes might be affected by CIC loss in different contexts, we performed a multi-gene list functional enrichment analysis of genes identified as being DE in our cell line models and in primary samples (Figure 5A). Functional terms enriched for DE genes were similar to those seen in the cell line models (supplementary material, Table S5A), with a smaller proportion of clusters (5/40 versus 9/40) relating to CNS development (supplementary material, Table S9A). Notably, however, the majority of these CNS development-related terms were significantly enriched in all four contexts, including *CIC*-deficient STAD samples. Once again, clusters of terms related to the development of other organs and systems were present (i.e. vasculature and heart, muscle, bone, and female sexual development). Interestingly, terms related to the epithelial–mesenchymal transition (EMT) and the cellular response to hypoxia, both of which have been associated with invasiveness and treatment resistance in glioma [49–51], were also significantly enriched, along with additional terms related to mesenchymal development and angiogenesis. Disruptions in WNT–β-catenin signalling and EMT also complement the apparent increase in cell motility conferred by loss of CIC [35].

Table 3. Number of putative CIC binding sites identified in the promoter regions of high-confidence candidate target genes.

Gene	Entrez ID	No. of putative binding sites
MATN2	4147	19
ETV4	2118	15
SLC13A3	64849	14
SPRED2	200734	13
TPD52L1	7164	12
PLS3	5358	9
PTPN9	5780	8
ZNF219	51222	8
ETV1	2115	8
LPAR1	1902	8
DUSP6	1848	7
SHC4	399694	7
ETV5	2119	7
SCN9A	6335	6
ENPP2	5168	6
NKAIN2	154215	6
SPRY4	81848	5
BTBD3	22903	5
EPHA2	1969	4
GLIPR1	11010	4
PRPS1	5631	4
SHC3	53358	4
GPM6B	2824	3
DUSP4	1846	3
SPRED1	161742	3
EDIL3	10085	2
FOSL1	8061	2
RAB31	11031	1
COL2A1	1280	1
CCND1	595	1
NUDT7	283927	1
GPR3	2827	1
GNG11	2791	1
EVI2A	2123	1
PLPPR5	163404	1
CRABP1	1381	1
RAET1E	135250	1
MFSD2A	84879	0
TRAPPC9	83696	0

Genes overexpressed in *CIC*-deficient samples showed enrichment of oncogenic signatures including gene sets that have been shown to be overexpressed upon activation of KRAS, EGFR, MEK, RAF, ERBB2, SRC, STK33, and CCND1 (Figure 5B; supplementary material, Table S9B). Hallmark gene sets related to upregulated KRAS signalling, hypoxia, and the p53 pathway were also significantly enriched. Consistent with these results, genes with reduced expression in *CIC*-deficient samples were enriched for genes that have been shown to have reduced expression upon activation of KRAS, RAF, MEK, or CCND1, or upon downregulation of RB (supplementary material, Table S9C).

Taken together, these results indicate that the gene expression differences seen in *CIC*-deficient samples are representative of gene expression dysregulation events frequently seen in various malignancies. They also show that, although the transcriptional consequences of CIC loss are, to some degree, context-dependent (supplementary material, Figure S6A), the functional consequences

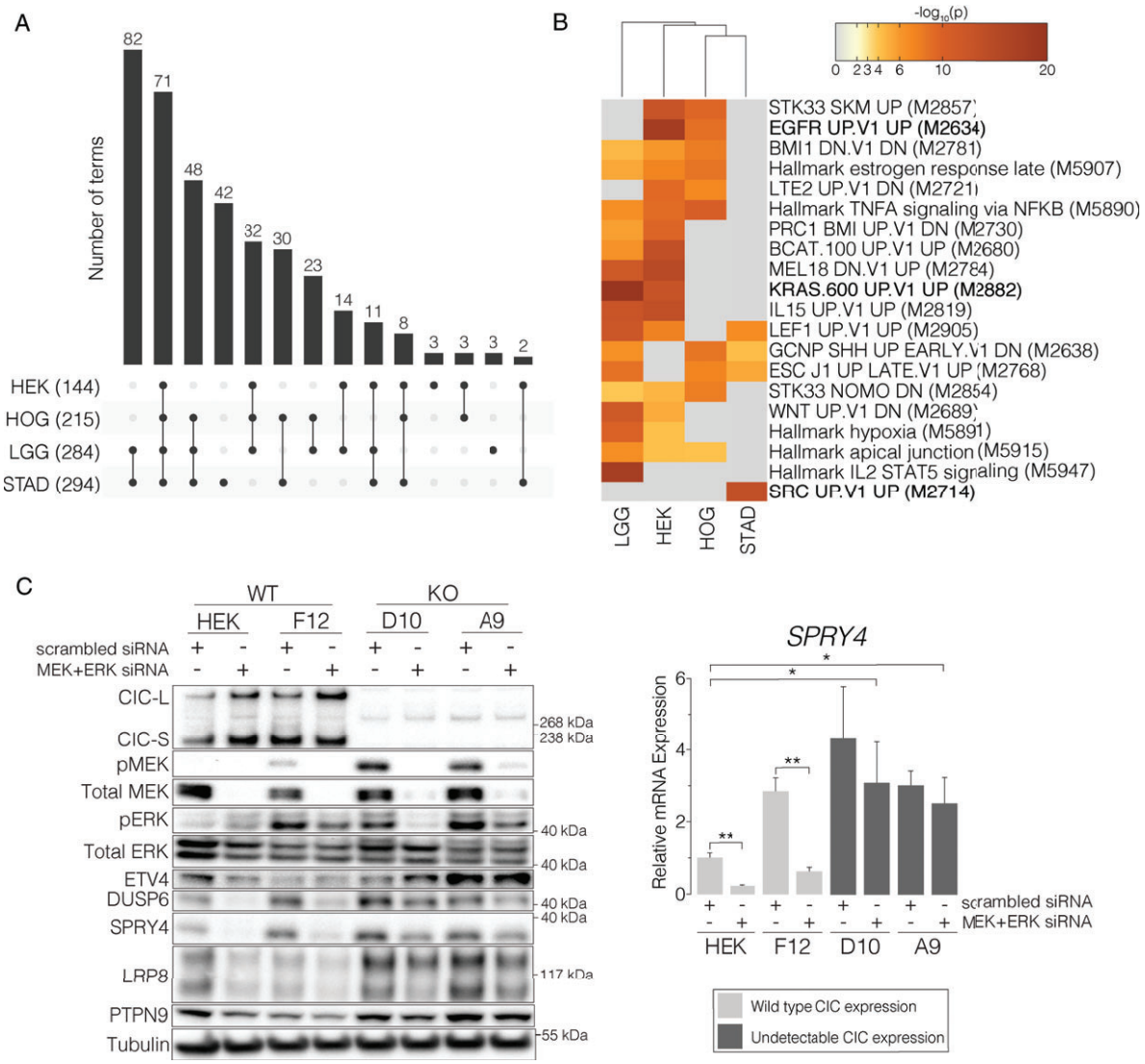

Figure 5. Gene expression differences associated with loss of CIC overlap with those associated with activation of MAPK signalling. (A) UpSet plot showing overlap of GO Biological Process terms significantly enriched for DE genes identified in the four contexts studied (Table S9A). Numbers in parentheses on the *x*-axis indicate the number of terms enriched for DE genes identified in each context, and numbers above bar plots indicate the number of terms in each overlap displayed below. (B) The most enriched terms from the top 10 clusters of Hallmark gene sets and Oncogenic signatures enriched for genes that show overexpression upon loss of CIC (Table S9B). Term IDs from MSigDB are shown. Terms related to MAPK signalling are in bold. (C) Left: representative western blots of *CIC*^{WT} and *CIC*^{KO} cell lines treated with a 'scrambled' non-targeting control siRNA or *MEK*-specific and *ERK*-specific siRNAs. Tubulin was used as a loading control, and a representative blot is shown. Right: quantification for SPRY4, shown as mean expression relative to HEK + scr siRNA. Additional quantifications are shown in supplementary material, Figure S6B. Error bars: standard error of the mean over three independent experiments. $^{*}p < 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$ (two-sided Student's *t*-test).

of CIC loss appear to be similar across biologically distinct contexts (Figure 5A). This is consistent with the notion that *CIC* mutations play an oncogenic role and can do so beyond the context of LGG.

Loss of CIC is associated with a MEK activation transcriptional signature

As noted, analyses of differential gene expression in *CIC*-deficient cell line models and primary samples indicated that loss of CIC is associated with dysregulation of the MAPK signalling cascade. Indeed, several of the high-confidence candidate target genes (*ETV1/4/5*, *DUSP4/6*, *SPRY4*, *SPRED2*, *GPR3*, *PTPN9*, and *LRP8*) have previously been identified as members of MEK [52,53] and/or RAS [54] activation signatures. This may indicate that the transcriptional dysregulation associated with CIC loss overlaps with activation of the MAPK signalling cascade.

To test this hypothesis, we used small interfering RNAs (siRNAs) to knock down *MEK1/2* (*MEK*) and *ERK2* (*ERK*) expression in *CIC*^{WT} and *CIC*^{KO} lines. The expression of candidate CIC target genes (*SPRY4*, *DUSP6*, *LRP8*, and *PTPN9*) was reduced in *CIC*^{WT} lines following *MEK*/*ERK* knockdown (Figure 5C; supplementary material, Figure S6B), consistent with results from previous studies [52–54]. These results are also consistent with studies showing that ERK activity leads

to CIC inhibition [11,18]; here, reduction of ERK activity could lead to relief of CIC inhibition, and thus to transcriptional repression of CIC target genes. Conversely, the expression of these target genes in *CIC*^{KO} lines following *MEK*/*ERK* siRNA treatment is decreased to a lesser extent, indicating that active CIC is at least partially required to transduce changes in MEK/ERK activity. Furthermore, *MEK*/*ERK* siRNA treatment is generally unable to 'rescue' the expression of candidate target genes. Similar results were obtained following treatment with a MEK inhibitor (supplementary material, Figure S6C). Thus, loss of CIC leads to aberrant overexpression of downstream MAPK targets in the absence of other common MAPK-activating mutations, indicating that it may present a novel mechanism for dysregulation of this common oncogenic pathway.

Discussion

Here, we explored CIC's transcriptional network in novel isogenic cell line models and in two biologically distinct cancer types. We identified 39 high-confidence candidate targets of CIC transcriptional regulation, including the established targets *ETV1*, *ETV4*, and *ETV5* [10,11,21,34]. We showed that this set of 39 genes appeared to be enriched for direct targets of CIC transcriptional regulation, and CIC binding in the promoter region of seven genes was confirmed by targeted ChIP-qPCR analysis. Interestingly, our results indicate that CIC missense mutants may retain some repressive activity. While this study was focused on truncating *CIC* mutations within type I LGGs, further analyses exploring the transcriptional programmes associated with *CIC* missense mutations may further inform on the potential role of this class of mutations.

This study is also the first to report an extensive list of candidate targets of CIC transcriptional regulation in STADs. A comparison of DE genes identified in biologically distinct contexts revealed that, although only *ETV4* was common to all contexts studied, similar biological processes and gene families appeared to be consistently affected. For instance, we observed several members of the *PCDH* gene family showing decreased expression in *CIC*-deficient samples. Reduced expression of several *PCDH* genes has been implicated in both low-grade and high-grade gliomas, including *PCDHGA11* [55], *PCDH10* [56], and *PCDH9* [57–59]. Similarly, hypermethylation and associated decreased expression of *PCDH10* [60–62], *PCDH8* [63] and *PDCH17* [64,65] have been associated with poor prognosis in gastric cancers. Thus, loss of CIC may affect cell adhesion processes through gene expression dysregulation, which is consistent with a recent report showing that loss of CIC in lung cancer cells leads to increased metastatic potential through elevated expression of ETV4 and matrix metalloproteinase-24 (MMP24) [35]. Other common pathways included the development of several tissue types, indicating that CIC may be more extensively involved in human development than currently appreciated.

We also observed an enrichment of known RTK–MAPK pathway regulators within DE genes, consistent with the notion that CIC may function in one or more feedback loop(s) to regulate MAPK signalling, as previously suggested [10,34]. Functional enrichment analyses also indicated that gene expression changes that occur upon loss of CIC significantly overlap with those that occur in response to increased MAPK signalling. We showed that MEK/ERK inhibition was able to reduce the expression of targets in *CIC*^{WT} lines, but less so or not at all in *CIC*^{KO} lines, indicating that CIC is needed, at least in part, to transduce signals from upstream members of the MAPK signalling pathway. Our results, combined with the observation that *CIC* mutations rarely co-occur with other activating alterations in this pathway [8], indicate that loss of CIC may provide a novel mechanism for activation of downstream members of the MAPK signalling cascade. These results are consistent with recent reports showing that loss of CIC imparts resistance to MAPK and EGFR inhibitors in various cancer-derived cell lines with activating mutations in upstream members of the pathway, including KRAS, NRAS, BRAF, and EGFR [23,24]. Although these reports show that increased expression of *ETV1/4/5* contributes to this phenotype, the additional CIC targets identified in our study may also play a role in this response. Our results thus expand on the potential roles of *CIC* mutations in malignancy, and may provide new insights into the possible mechanisms underlying phenotypic responses recently associated with CIC loss, such as shorter times to recurrence, increased metastatic potential, and resistance to MAPK inhibitors [9,10,23,24,35].

Acknowledgements

We thank Dr Gregory Cairncross, Dr Jennifer Chan and members of the Marra laboratory for helpful discussions. MAM gratefully acknowledges the support of the BC Cancer Agency, the BC Cancer Foundation, the Canada Research Chairs program, the University of British Columbia, and the Canadian Institutes of Health Research (CIHR; FDN-143288). We are especially grateful to Ms Donna Anderson for her generosity in supporting this project. We thank Dr G. Dawson (The University of Chicago, Illinois, USA) for providing us with the HOG cell line. VGL is supported by a Canadian Institutes of Health Research (CIHR) Canada Graduate Scholarship Master's Award, a Killam Doctoral Scholarship, and a CIHR Vanier Canada Graduate Scholarship. SY is supported by a VCHRI Mentored Scientist Award and acknowledges the ongoing support of BrainCare BC. The results published here are in part based upon data generated by The Cancer Genome Atlas managed by the NCI and NHGRI. Information about TCGA can be found at http://cancergenome.nih.gov.

Author contributions statement

VGL, MF, SC, MAM: conceived and designed the study; VGL: performed bioinformatics analyses and, along with MAM, wrote the manuscript; MF: developed the ZFN *CIC* knockout cell line; JS, SYC, AL, SC: developed the CRISPR/Cas9 *CIC* knockout cell lines; JS, SYC: performed most cell line-based experiments; MAM: supervised the project; SC, SY: provided further guidance. All authors participated in discussions regarding the experiments and results, and reviewed and approved the manuscript.

References

1. Suzuki H, Aoki K, Chiba K, *et al.* Mutational landscape and clonal architecture in grade II and III gliomas. *Nat Genet* 2015; **47:** 458–468.
2. Weller M, Weber RG, Willscher E, *et al.* Molecular classification of diffuse cerebral WHO grade II/III gliomas using genome- and transcriptome-wide profiling improves stratification of prognostically distinct patient groups. *Acta Neuropathol* 2015; **129:** 679–693.
3. Dubbink HJ, Atmodimedjo PN, Kros JM, *et al.* Molecular classification of anaplastic oligodendroglioma using next-generation sequencing: a report of the prospective randomized EORTC Brain Tumor Group 26951 phase III trial. *Neuro Oncol* 2015; **18**: 388–400.
4. Louis DN, Ohgaki H, Wiestler OD, *et al* (eds). *World Health Organization Classification of Tumours of the Central Nervous System (4th edn)*. IARC: Lyon, 2016.
5. Bettegowda C, Agrawal N, Jiao Y, *et al.* Mutations in CIC and FUBP1 contribute to human oligodendroglioma. *Science* 2011; **333:** 1453–1455.
6. Yip S, Butterfield YS, Morozova O, *et al.* Concurrent CIC mutations, IDH mutations, and 1p/19q loss distinguish oligodendrogliomas from other cancers. *J Pathol* 2012; **226:** 7–16.
7. Sahm F, Koelsche C, Meyer J, *et al.* CIC and FUBP1 mutations in oligodendrogliomas, oligoastrocytomas and astrocytomas. *Acta Neuropathol* 2012; **123:** 853–860.
8. The Cancer Genome Atlas Research Network. Comprehensive, integrative genomic analysis of diffuse lower-grade gliomas. *N Engl J Med* 2015; **372:** 2481–2498.
9. Chan AK, Pang JC, Chung NY, *et al.* Loss of CIC and FUBP1 expressions are potential markers of shorter time to recurrence in oligodendroglial tumors. *Mod Pathol* 2014; **27:** 332–342.
10. Gleize V, Alentorn A, Connen de Kerillis L, *et al.* CIC inactivating mutations identify aggressive subset of 1p19q codeleted gliomas. *Ann Neurol* 2015; **78**: 355–374.
11. Jimenez G, Shvartsman SY, Paroush Z. The Capicua repressor – a general sensor of RTK signaling in development and disease. *J Cell Sci* 2012; **125:** 1383–1391.
12. Jin Y, Ha N, Fores M, *et al.* EGFR/Ras signaling controls Drosophila intestinal stem cell proliferation via Capicua-regulated genes. *PLoS Genet* 2015; **11:** e1005634.
13. Yang L, Paul S, Trieu KG, *et al.* Minibrain and Wings apart control organ growth and tissue patterning through down-regulation of Capicua. *Proc Natl Acad Sci U S A* 2016; **113:** 10583–10588.
14. Astigarraga S, Grossman R, Diaz-Delfin J, *et al.* A MAPK docking site is critical for downregulation of Capicua by Torso and EGFR RTK signaling. *EMBO J* 2007; **26:** 668–677.
15. Jimenez G, Guichet A, Ephrussi A, *et al.* Relief of gene repression by torso RTK signaling: role of capicua in Drosophila terminal and dorsoventral patterning. *Genes Dev* 2000; **14:** 224–231.
16. Lee CJ, Chan WI, Cheung M, *et al.* CIC, a member of a novel subfamily of the HMG-box superfamily, is transiently expressed in developing granule neurons. *Brain Res Mol Brain Res* 2002; **106:** 151–156.
17. Chen T, Zhou L, Yuan Y, *et al.* Characterization of Bbx, a member of a novel subfamily of the HMG-box superfamily together with Cic. *Dev Genes Evol* 2014; **224:** 261–268.
18. Futran AS, Kyin S, Shvartsman SY, *et al.* Mapping the binding interface of ERK and transcriptional repressor Capicua using photocrosslinking. *Proc Natl Acad Sci U S A* 2015; **112:** 8590–8595.
19. Kawamura-Saito M, Yamazaki Y, Kaneko K, *et al.* Fusion between CIC and DUX4 up-regulates PEA3 family genes in Ewing-like sarcomas with t(4;19)(q35;q13) translocation. *Hum Mol Genet* 2006; **15:** 2125–2137.
20. Dissanayake K, Toth R, Blakey J, *et al.* ERK/p90(RSK)/14-3-3 signalling has an impact on expression of PEA3 Ets transcription factors via the transcriptional repressor capicua. *Biochem J* 2011; **433:** 515–525.
21. Specht K, Sung YS, Zhang L, *et al.* Distinct transcriptional signature and immunoprofile of CIC–DUX4 fusion-positive round cell tumors compared to EWSR1-rearranged Ewing sarcomas: further evidence toward distinct pathologic entities. *Genes Chromosomes Cancer* 2014; **53:** 622–633.
22. Oh S, Shin S, Janknecht R. ETV1, 4 and 5: an oncogenic subfamily of ETS transcription factors. *Biochim Biophys Acta* **2012**; 1826: 1–12.
23. Wang B, Krall EB, Aguirre AJ, *et al.* ATXN1L, CIC, and ETS transcription factors modulate sensitivity to MAPK pathway inhibition. *Cell Rep* 2017; **18:** 1543–1557.
24. Liao S, Davoli T, Leng Y, *et al.* A genetic interaction analysis identifies cancer drivers that modify EGFR dependency. *Genes Dev* 2017; **31**: 184–196.
25. Carvalho BS, Irizarry RA. A framework for oligonucleotide microarray preprocessing. *Bioinformatics* 2010; **26:** 2363–2367.
26. Dalman MR, Deeter A, Nimishakavi G, *et al.* Fold change and p-value cutoffs significantly alter microarray interpretations. *BMC Bioinformatics* 2012; **13**(suppl 2): S11.
27. Eisenreich S, Abou-El-Ardat K, Szafranski K, *et al.* Novel CIC point mutations and an exon-spanning, homozygous deletion identified in oligodendroglial tumors by a comprehensive genomic approach including transcriptome sequencing. *PLoS One* 2013; **8:** e76623.
28. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014; **15:** 550.
29. Lombardo A, Genovese P, Beausejour CM, *et al.* Gene editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery. *Nat Biotechnol* 2007; **25:** 1298–1306.
30. Mali P, Yang L, Esvelt KM, *et al.* RNA-guided human genome engineering via Cas9. *Science* 2013; **339:** 823–826.
31. Cong L, Ran FA, Cox D, *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* 2013; **339:** 819–823.
32. Post GR, Dawson G. Characterization of a cell line derived from a human oligodendroglioma. *Mol Chem Neuropathol* 1992; **16:** 303–317.
33. Chittaranjan S, Chan S, Yang C, *et al.* Mutations in CIC and IDH1 cooperatively regulate 2-hydroxyglutarate levels and cell clonogenicity. *Oncotarget* 2014; **5:** 7960–7979.
34. Padul V, Epari S, Moiyadi A, *et al.* ETV/Pea3 family transcription factor-encoding genes are overexpressed in CIC-mutant oligodendrogliomas. *Genes Chromosomes Cancer* 2015; **54**: 725–733.
35. Okimoto RA, Breitenbuecher F, Olivas VR, *et al*. Inactivation of Capicua drives cancer metastasis. *Nat Genet* 2017; **49**: 87–96.
36. Choi N, Park J, Lee JS, *et al.* miR-93/miR-106b/miR-375-CIC-CRABP1: a novel regulatory axis in prostate cancer progression. *Oncotarget* 2015; **6**: 23533–23547.

37. Roch F, Jimenez G, Casanova J. EGFR signalling inhibits Capicua-dependent repression during specification of Drosophila wing veins. *Development* 2002; **129:** 993–1002.
38. Atkey MR, Lachance JF, Walczak M, *et al.* Capicua regulates follicle cell fate in the Drosophila ovary through repression of mirror. *Development* 2006; **133:** 2115–2123.
39. Ajuria L, Nieva C, Winkler C, *et al.* Capicua DNA-binding sites are general response elements for RTK signaling in Drosophila. *Development* 2011; **138:** 915–924.
40. Lee Y, Fryer JD, Kang H, *et al.* ATXN1 protein family and CIC regulate extracellular matrix remodeling and lung alveolarization. *Dev Cell* 2011; **21:** 746–757.
41. Kim E, Park S, Choi N, *et al.* Deficiency of Capicua disrupts bile acid homeostasis. *Sci Rep* 2015; **5:** 8272.
42. Ceccarelli M, Barthel FP, Malta TM, *et al.* Molecular profiling reveals biologically discrete subsets and pathways of progression in diffuse glioma. *Cell* 2016; **164:** 550–563.
43. Ohba S, Mukherjee J, See WL, *et al.* Mutant IDH1-driven cellular transformation increases RAD51-mediated homologous recombination and temozolomide resistance. *Cancer Res* 2014; **74**: 1836–4844.
44. Fores M, Ajuria L, Samper N, *et al.* Origins of context-dependent gene repression by capicua. *PLoS Genet* 2015; **11:** e1004902.
45. Moller E, Hornick JL, Magnusson L, *et al.* FUS-CREB3L2/L1-positive sarcomas show a specific gene expression profile with upregulation of CD24 and FOXL1. *Clin Cancer Res* 2011; **17:** 2646–2656.
46. Kim Y, Iagovitina A, Ishihara K, *et al.* Context-dependent transcriptional interpretation of mitogen activated protein kinase signaling in the Drosophila embryo. *Chaos* 2013; **23:** 025105.
47. Graham C, Chilton-MacNeill S, Zielenska M, *et al.* The CIC–DUX4 fusion transcript is present in a subgroup of pediatric primitive round cell sarcomas. *Hum Pathol* 2012; **43:** 180–189.
48. The Cancer Genome Atlas Research Network. Comprehensive molecular characterization of gastric adenocarcinoma. *Nature* 2014; **513:** 202–209.
49. Karsy M, Guan J, Jensen R, *et al.* The impact of hypoxia and mesenchymal transition on glioblastoma pathogenesis and cancer stem cells regulation. *World Neurosurg* 2016; **88:** 222–236.
50. Iser IC, Pereira MB, Lenz G, *et al.* The epithelial-to-mesenchymal transition-like process in glioblastoma: an updated systematic review and in silico investigation. *Med Res Rev* 2017; **37:** 271–313.
51. Oliver L, Olivier C, Marhuenda FB, *et al.* Hypoxia and the malignant glioma microenvironment: regulation and implications for therapy. *Curr Mol Pharmacol* 2009; **2:** 263–284.
52. Nazarian R, Shi H, Wang Q, *et al.* Melanomas acquire resistance to B-RAF(V600E) inhibition by RTK or N-RAS upregulation. *Nature* 2010; **468:** 973–977.
53. Ambrosini G, Pratilas CA, Qin LX, *et al.* Identification of unique MEK-dependent genes in GNAQ mutant uveal melanoma involved in cell growth, tumor cell invasion, and MEK resistance. *Clin Cancer Res* 2012; **18:** 3552–3561.
54. Loboda A, Nebozhyn M, Klinghoffer R, *et al.* A gene expression signature of RAS pathway dependence predicts response to PI3K and RAS pathway inhibitors and expands the population of RAS pathway activated tumors. *BMC Med Genomics* 2010; **3:** 26.
55. Waha A, Guntner S, Huang TH, *et al.* Epigenetic silencing of the protocadherin family member PCDH-gamma-A11 in astrocytomas. *Neoplasia* 2005; **7:** 193–199.
56. Echizen K, Nakada M, Hayashi T, *et al.* PCDH10 is required for the tumorigenicity of glioblastoma cells. *Biochem Biophys Res Commun* 2014; **444:** 13–18.
57. de Tayrac M, Etcheverry A, Aubry M, *et al.* Integrative genome-wide analysis reveals a robust genomic glioblastoma signature associated with copy number driving changes in gene expression. *Genes Chromosomes Cancer* 2009; **48:** 55–68.
58. Wang C, Yu G, Liu J, *et al.* Downregulation of PCDH9 predicts prognosis for patients with glioma. *J Clin Neurosci* 2012; **19:** 541–545.
59. Wang C, Tao B, Li S, *et al.* Characterizing the role of PCDH9 in the regulation of glioma cell apoptosis and invasion. *J Mol Neurosci* 2014; **52:** 250–260.
60. Yu J, Cheng YY, Tao Q, *et al.* Methylation of protocadherin 10, a novel tumor suppressor, is associated with poor prognosis in patients with gastric cancer. *Gastroenterology* 2009; **136:** 640–651.
61. Li Z, Chim JC, Yang M, *et al.* Role of PCDH10 and its hypermethylation in human gastric cancer. *Biochim Biophys Acta* 2012; **1823:** 298–305.
62. Deng J, Liang H, Ying G, *et al.* Clinical significance of the methylated cytosine-phosphate-guanine sites of protocadherin-10 promoter for evaluating the prognosis of gastric cancer. *J Am Coll Surg* 2014; **219:** 904–913.
63. Zhang D, Zhao W, Liao X, *et al.* Frequent silencing of protocadherin 8 by promoter methylation, a candidate tumor suppressor for human gastric cancer. *Oncol Rep* 2012; **28:** 1785–1791.
64. Yang Y, Liu J, Li X, *et al.* PCDH17 gene promoter demethylation and cell cycle arrest by genistein in gastric cancer. *Histol Histopathol* 2012; **27:** 217–224.
65. Hu X, Sui X, Li L, *et al.* Protocadherin 17 acts as a tumour suppressor inducing tumour cell apoptosis and autophagy, and is frequently methylated in gastric and colorectal cancers. *J Pathol* 2013; **229:** 62–73.
*66. Lawrence M, Huber W, Pages H, *et al.* Software for computing and annotating genomic ranges. *PLoS Comput Biol* 2013; **9:** e1003118.
*67. Zhu LJ, Christensen RG, Kazemian M, *et al.* FlyFactorSurvey: a database of Drosophila transcription factor binding specificities determined using the bacterial one-hybrid system. *Nucleic Acids Res* 2011; **39:** D111–D117.
*68. Myers S, Bottolo L, Freeman C, *et al.* A fine-scale map of recombination rates and hotspots across the human genome. *Science* 2005; **310:** 321–324.
*69. Tripathi S, Pohl MO, Zhou Y, *et al.* Meta- and orthogonal integration of Influenza 'OMICs' data defines a role for UBR4 in virus budding. *Cell Host Microbe* 2015; **18:** 723–735.

*Cited only in supplementary material.

SUPPLEMENTARY MATERIAL ONLINE

Supplementary materials and methods

Supplementary figure legends

Figure S1. *CIC* expression in Type I LGGs with intact *CIC* (WT) or truncating *CIC* mutations (Mut)

Figure S2. Generation of *CIC* knockout cell lines

Figure S3. *CIC* missense mutants retain repressive activity

Figure S4. ETV4 shows increased protein expression in CIC^{KO} cell lines