CIZ1-F, an alternatively spliced variant of the DNA replication protein CIZ1 with distinct expression and localisation, is overrepresented in early stage common solid tumours

Dorian R.A. Swarts[1], Emma R. Stewart[2], Gillian S. Higgins[3], Dawn Coverley[4]

Department of Biology, The University of York, York, United Kingdom

[1]E-mail: draswarts@hotmail.com; ORCID ID: https://orcid.org/0000-0002-7927-7613

[2]E-mail: ers518@york.ac.uk; ORCID ID: https://orcid.org/0000-0003-0189-9348

[3]E-mail: gillian.higgins@york.ac.uk

[4]E-mail: dawn.coverley@york.ac.uk; ORCID ID: https://orcid.org/0000-0001-8262-7023

Address correspondence to:

Dawn Coverley, PhD

E-mail: dawn.coverley@york.ac.uk

Tel. +44-(0)1904-328569

Wentworth Way

York YO10 5DD

United Kingdom

Word count Introduction-Discussion: 4722

Abstract

CIZ1 promotes cyclin-dependent DNA replication and resides in sub-nuclear foci that are part of the protein nuclear matrix (NM), and in RNA assemblies that are enriched at the inactive X chromosome (Xi) in female cells. It is subjected to alternative splicing, with specific variants implicated in adult and paediatric cancers. CIZ1-F is characterized by a frame shift that results from splicing exons 8-12 leading to inclusion of a short alternative reading frame (ARF), excluding the previously characterized C-terminal NM anchor domain. Here, we apply a set of novel variant-selective molecular tools targeted to the ARF to profile the expression of CIZ1-F at both transcript and protein levels, with focus on its relationship with the RNA-dependent and -independent fractions of the NM. Unlike full-length CIZ1, CIZ1-F does not accumulate at Xi, though like full-length CIZ1 it does resist extraction with DNase. Notably, CIZ1-F is sensitive to RNase identifying it as part of the RNA-fraction of the NM. In quiescent cells *CIZ1*-F transcript expression is suppressed and CIZ1-F protein is excluded from the nucleus, with re-expression not observed until the second cell cycle after exit from quiescence. Importantly, *CIZ1*-F is over-expressed in common solid tumours including colon and breast, pronounced in early stage but not highly-proliferative late stage tumours. Moreover, expression was significantly higher in hormone receptor negative breast tumours than receptor positive tumours. Together these data show that CIZ1-F is expressed in proliferating cells in an unusual cell cycle-dependent manner, and suggest that it may have potential as a tumour biomarker.

Word-count: 250 [max. 250]

Keywords: CDKN1A-interacting zinc finger protein 1, nuclear matrix, alternative splicing.

Disclosure statement

All authors declare that they have no competing interests.

Funding details

This work was funded by a project grant from Yorkshire Cancer Research (YCR grant no Y258 to DC) and BBSRC doctoral training funds to ES (White Rose DTP grant no BB/M011151/1).

Introduction

CDKN1A interacting zinc finger protein 1 (CIZ1) is subject to extensive alternative splicing to yield multiple transcript variants.[1-5] Functional analysis of discrete protein fragments has identified domains and regulatory sites in the N-terminal half of the protein that are involved in DNA replication[6, 7] and interaction with cell cycle regulators[7] (**Fig. 1A, B**). Separate sequences in the C-terminus of CIZ1 support interaction with nuclear structures, identifying a salt and DNase-resistant nuclear matrix (NM) interaction domain,[8] which may restrict its function to specific sub-nuclear locations. Recently, we and others showed that CIZ1 localises to the inactive X chromosome (Xi) and is required to ensure retention of Xist long non-coding RNA (lncRNA) at Xi in fibroblasts.[9, 10] Alternative splicing affects both the replication domain (RD) and NM anchor domain (AD) of CIZ1, and several splicing events have been implicated in cancer, including mutation driven exclusion of exon 4 (CIZ1-Δ4) in Ewing's Tumour cell lines[2] and exclusion of 24 nucleotides of exon 14 in lung cancer (CIZ1-B).[5] A unique splice-junction epitope from CIZ1-B is a circulating biomarker for patients with small cell and non-small cell lung cancer.[5, 11] CIZ1 has also been reported to be involved in the development of breast cancer,[12] the most common type of cancer in women.[13] Alternative splicing of CIZ1 is implicated in other chronic diseases including upregulation of a form in which part of exon 8 is excluded (CIZ1-S) in Alzheimer's disease,[3] and in cervical dystonia where mutations in an exonic splicing enhancer may affect splicing patterns.[14]

Despite functional studies that link full-length murine CIZ1 and embryonic variant murine CIZ1 (ECIZ1) with DNA replication,[6, 7] little is known about the function of other variants. Alternative splicing of *CIZ1* may have important consequences for the cell and is known to affect the sub-nuclear localisation of the protein (**Supplementary Table 1**). For example, exclusion of exon 4 changes the sub-nuclear distribution of CIZ1 from focal to

diffuse.[2] Thus alternative splicing of *CIZ1* may influence the spatial organisation of DNA replication, and suggests that some CIZ1 splice variants have the potency to act as dominant-negative controllers of other variants.[2]

We previously designed an exon-junction microarray that identified a number of novel variants of CIZ1, including one common cancer-associated *CIZ1*-variant characterised by a continuous deletion of part of exon 8, exon 9-11 and part of exon 12, here referred to as CIZ1-F (**Fig. 1A**).[1] CIZ1-F was initially identified in Ewing's Tumour cell lines and found to be overexpressed in primary lung tumours but not matched normal tissues.[1] Here, we characterise CIZ1-F mRNA and protein expression patterns by exploiting a short unique peptide encoded by expression of an alternative reading frame (ARF) of exon 12-13. We show that CIZ1-F is specifically expressed in G1 phase following mitosis but not G1 phase following quiescence. Despite lacking the C-terminal NM-attachment domain[8] and several protein interaction sites that are important for replication function, CIZ1-F resists extraction with DNase1. It is however sensitive to RNase, which contrasts with the behaviour of full-length CIZ1. Most importantly, *CIZ1*-F is overexpressed in early stage primary human tumours and is associated with ER-negative status in breast tumours.

Results

Human *CIZ1*-F is characterised by an 1181 base pair deletion (c1038_2218del1181 in reference sequence NM_012127.2), which excludes part of exons 8 and 12 and all of exons 9-11 (**Fig. 1A**, **Supplementary Fig. 1A**). The exon 8-12 junction causes a frame-shift in the *CIZ1* reading frame, leading to a premature translational stop codon in exon 13, despite the presence of downstream exons in *CIZ1*-F mRNA (**Fig. 1A**). CIZ1-F protein therefore lacks

previously characterised functional domains, including the MH3-domain which anchors CIZ1 to the NM,[8] and cyclin-binding motifs through which CIZ1 interacts with cyclin A, and which are essential for its DNA replication activity (**Fig. 1B**).[6, 7] Other characterised sites of interaction with CDK2, p21^{cip1} and YAP are also excluded,[15-17] while interaction with CDC6,[18] ERα,[12] and dynein light chain[15] might be retained. The interaction sites of the CIZ1 binding factors PDRG1,[19] SH3BP4,[20] and TCF4[21] are not yet characterised. Importantly, the ARF created by the 8-12 junction encodes 37 amino acids of unique sequence not present in other forms of CIZ1, or in other proteins, and results in inclusion of two additional LXXLL oestrogen receptor (ER)-interaction motifs (**Fig. 1B, C**).[22] Thus *CIZ1*-F transcript appears to encode a 315 amino acid protein with a unique C-terminal end. To study *CIZ1*-F mRNA expression we designed a *CIZ1*-F junction-specific gene expression probe (**Supplementary Table 2**, **Supplementary Fig. 1B-C**). Initially, *CIZ1*-F mRNA levels were measured by qRT-PCR in a set of cell lines (**Supplementary Fig. 1D**), revealing variable levels, with high expression in bladder and cervical cancer lines EJ and HeLa, as well as in breast cancer cell lines MCF-7, BT474 and MDA-MB-231. Evaluation of cultured tumour-derived cell lines showed that *CIZ1*-F transcript is elevated in MCF7 breast cancer cells compared to normal MCF-10A breast cells (**Supplementary Fig. 1D**), and is in fact one of the main *CIZ1* PCR-products (**Supplementary Fig. 1A**), identifying MCF7 cells as a model cell line to study the function of CIZ1-F in cell proliferation. However, despite the reported functional interaction between CIZ1 and ERα,[12] we were unable to detect a reproducible effect of oestrogen on the expression of *CIZ1* or *CIZ1*-F transcript in MCF7 cells (**Supplementary Fig. 1E**).

CIZ1-F protein is part of the RNA-dependent NM

In order to confirm that *CIZ1*-F transcript encodes a protein we generated an affinity-purified anti-peptide antibody against sequences in the unique C-terminal ARF of predicted CIZ1-F

protein (**Fig. 1C**). Specificity of the antibody was confirmed using ectopically expressed EGFP-full-length CIZ1 and EGFP-CIZ1-F for western blot and immunofluorescence applications (**Supplementary Fig. 2A-B**). Further confirmation of the specificity of our CIZ1-F antibody was obtained using peptide blocking, which yielded two CIZ1-F-specific products of approximately 42kD and 71kD in total cell lysates (**Supplementary Fig. 2C**). The 42kD monomeric CIZ1-F product is larger than the predicted molecular weight of CIZ1-F (34kD) but in agreement with the mobility of EGFP-CIZ1-F after subtraction of EGFP (45kD; **Supplementary Fig. 2A**).

Since full-length CIZ1 localises to the Xi and interacts with Xist lncRNA, we first determined the relationship between CIZ1-F and the Xi in MCF-7 cells. Full-length CIZ1, detected by a monoclonal antibody to a C-terminal epitope, clearly localizes to the Xi in MCF-7 cells, though in line with reports that Xi in breast cancer-derived cells is a less discrete (compact) entity than in 'normal' cells[23] Xi staining is markedly irregular, and also accompanied by the nucleus wide signal that is typical of CIZ1.[9] In contrast, CIZ1-F shows a different pattern including both nuclear and cytoplasmic foci, but no enrichment at Xi (**Figure 2A-B**). We next examined the relationship of full-length CIZ1 and CIZ1-F with the NM. Full-length CIZ1 is part of the core protein NM (resistant to extraction with DNase and RNase), as evident by immunodetection of endogenous CIZ1 with RD-specific antibodies (**Fig. 2C-F**). Analysis of attachment to this core protein NM can be recapitulated using EGFP-full-length CIZ1 and derived fragments, allowing the domain responsible for attachment to the NM to be located at the C-terminal end, and classified as 'anchor domain' or AD.[8] AD is missing from CIZ1-F, which would therefore not be expected to resist extraction. Importantly, under identical conditions to those used to profile endogenous CIZ1, the extraction profile for CIZ1-F indicates attachment to the RNA component of the NM (RNA-protein NM) because it resists extraction with DNase, but not when used in

conjunction with RNase (**Fig. 2C-F**). Notably, direct comparison of immunofluorescence signal in nuclei at different stages in the extraction process shows that the majority of the CIZ1-F epitope in the nucleus is in fact masked by chromatin (**Fig. 2E**). The signal seen before extraction appears to be not the same population as that revealed when chromatin is removed, because pre-extraction signal is lost in mock treated nuclei and only revealed if DNase is also included in the reaction (**Fig. 2E**). Again, despite the reported association between ERα and CIZ1,[12] we did not observe co-localisation of CIZ1-F with nuclear ERα in immuno-detection experiments (**Supplementary Fig. 2D**).

In conclusion, CIZ1-F resists extraction of chromatin and is therefore a NM protein, however it is not part of the core protein NM like full-length CIZ1 as it can be fully extracted by digestion of nuclear RNA, in the presence or absence of co-extraction of DNA, identifying it as part of the RNA-protein NM (**Fig. 2G**). The differential extraction profiles of full-length CIZ1 and CIZ1-F epitopes suggest that they may have distinct functions within the NM.

CIZ1-F expression in cycling and quiescent cells

We next focused on CIZ1-F expression in cycling and quiescent cell populations, since CIZ1-F retains most regions involved in DNA-replication (**Fig. 1B**). In detergent-treated cell populations, endogeneous CIZ1-F protein expression patterns changed on entry to quiescence, switching from nuclear localisation in cycling cells, to nuclear exclusion with cytoplasmic signal at confluence, to nearly undetectable in quiescence (**Fig. 3A-B**). Importantly, when quiescent cell populations were subjected to full NM-extraction, all signal was lost when cells were treated with DNase, demonstrating that quiescent cell populations do not retain any CIZ1-F on the NM (**Fig. 3C**).

CIZ1-F transcript levels were also assessed in cycling, contact-inhibited and serum-deprived cells, and compared to amplicons in *CIZ1*-RD and *CIZ1*-AD (**Fig. 3D-E**). Overall,

CIZ1-F transcript is approximately 50-fold lower than 'total' *CIZ1* mRNA in cycling cells, which is broadly consistent with the frequency of expressed sequence tags in NCBI UniGene (accessed 03/03/2018), where 11 of 875 sequences (all from cancers) are *CIZ1*-F. *CIZ1*-RD and *CIZ1*-AD levels remain relatively stable during proliferative growth, and slightly increase during growth rate decline caused by serum starvation. However, *CIZ1*-F mRNA expression fell dramatically as proliferation rate dropped (**Fig. 3D; Supplementary Fig. 3A**), and recovered again upon addition of fresh serum. In contrast, contact-inhibited normal foetal lung fibroblast MRC5 cells did not resume proliferation or restore CIZ1-F expression when boosted with serum (**Fig. 3E**). This suggests that stable contact-induced growth arrest suppresses expression of CIZ1-F. In the same MCF-7 cell populations we observed additional effects on *CIZ1* alternative splicing, most notably increased *CIZ1*-S (**Supplementary Fig. 3A**) an isoform reported to be upregulated in Alzheimer's disease (**Supplementary Table 1**), and changes in *CIZ1*-Δ4 (**Supplementary Fig. 3B**).[2, 3] Together the data show that CIZ1-F mRNA and protein are expressed in proliferating cells and inhibited by both serum-deprivation-induced proliferation arrest (MCF-7), and contact-inhibition-induced proliferation arrest (MRC5).

CIZ1-F mRNA is expressed in G1 following mitosis (M-G1) but not G1 following quiescence (Q-G1)

To understand better the role of *CIZ1*-F in proliferation we analysed its cell cycle expression profile. In order to generate synchronised populations, serum-deprived MCF-7 cells were released back into cycle in the presence of fresh serum and the nucleotide analogue EdU (**Fig. 4A**). The frequency of EdU-positive cells indicates that the population passes through S phase between 10 and 20 h after exiting quiescence, peaking at 16 h (49% of nuclei; **Fig. 4A**), and this is preceded as expected, by an increase in expression of cyclin E mRNA. However,

CIZ1-F transcript does not begin to increase until after the majority of cells have passed through S-phase, recovering only at 31 h (**Fig. 4A**). In contrast, *CIZ1* AD and RD expression remains relatively stable. When MCF-7 cells were exposed to the DNA synthesis inhibitors aphidicolin or thymidine, or arrested in mitosis by nocodazole upon release from quiescence, *CIZ1*-F expression was strongly suppressed compared to uninhibited cells (**Fig. 4B**). Similar results were obtained with MRC5 cell populations treated with thymidine or nocodazole (**Fig. 4C**). Thus *CIZ1*-F mRNA levels do not accumulate in either S-phase or mitosis and *CIZ1*-F is not expressed in Q-G1. However, when MCF7 cells were released from arrest with nocodazole in the M-phase following quiescence, *CIZ1*-F mRNA was expressed within 2 h, which is at least 4 h before cyclin E rises (**Fig. 4D**). Thus, *CIZ1*-F mRNA is expressed in early G1 phase in the second cell cycle following release from quiescence, or M-G1 (**Fig. 4E**).

CIZ1-F mRNA expression in primary tumours and cancer cell lines

To confirm that *CIZ1*-F is overrepresented in cancer tissue, and to extend our analysis of lung tumours,[1] we analysed *CIZ1* mRNA in colon and breast cancer. Almost without exception colon cancer samples expressed elevated *CIZ1*-F compared to matched normal control tissue ($P < 0.001$, Related-Samples Wilcoxon Signed Rank Test; **Fig. 5A**; see **Supplementary Fig. 4** for individual sample data), and surprisingly the highest levels were in early stage tumours ($P = 0.017$ for early versus late stage, Mann-Whitney U-test). In contrast *CIZ1*-RD does not increase as dramatically in early stages, nor diminish significantly at later stages ($P = 0.44$ for the same comparison, Mann-Whitney U-test; **Fig. 5A**). Thus the data indicate upregulation of *CIZ1*-F specifically in early stage colon cancer. Similar results were obtained with breast tumours with higher *CIZ1*-F expression in tumour tissue compared to normal breast tissues ($P = 0.016$, Mann-Whitney U-test), whereas expression of *CIZ1*-RD was not significantly

different (**Fig. 5B**; see **Supplementary Fig. 5** for individual data).[1] Again *CIZ1*-F levels were most elevated in early stage tumours ($P = 0.006$ when compared to normal tissues, Mann-Whitney U-test). Thus the profile of *CIZ1*-F expression across tumour stage is similar for both breast cancer and colon cancer, and further preliminary analysis of lung and urinary bladder cancers returned similar trends (**Supplementary Fig. 6**), suggesting that the stage profile may be common to other tumour types.

When grouped by tumour grade rather than stage, the greatest elevation was at lowest grade for colon cancer (**Fig. 5C**), but for breast cancer *CIZ1*-F levels increased with grade ($P = 0.029$, Kruskall-Wallis Test; **Fig. 5D**). This is not surprising however since, in this series, most low-grade breast cancers were high-stage, and vice versa (see **Fig. 5D**).

To evaluate the relationship with proliferation, the transcript level of the proliferation marker *MKI67* was measured in the same breast and colon samples (**Fig. 5**). For colon, although *MKI67* and *CIZ1*-F levels correlated closely when the matched normal tissues were included in the analysis ($P = 0.0020$; Pearson Correlation), no correlation was found between levels across the cancer samples only, whereas *CIZ1*-RD was strongly correlated to *MKI67* levels by both analyses ($P \leq 0.00043$; Pearson Correlation). Although less obvious for breast tumours, again no significant correlation across the full set for *CIZ1*-F and *MKI67* could be identified ($P = 0.058$; Pearson Correlation), compared to correlation with RD ($P \leq 0.0050$; Pearson Correlation). When performing linear regression with both *CIZ1*-RD and *CIZ1*-F in colon or breast samples, only *CIZ1*-RD significantly relates to *MKI67* expression in all analyses ($P \leq 0.0050$). Thus, unlike *CIZ1*-RD, *CIZ1*-F does not mirror expression of *MKI67* and therefore does not seem to have a simple relationship with cell proliferation.

Interestingly, *CIZ1*-F levels correlated with hormone receptor status in breast tumours. ER and PR negative tumours have significantly higher *CIZ1*-F transcript levels than ER or progesterone receptor (PR) positive tumours ($P = 0.005$, Mann-Whitney U-test; **Fig.**

6A-B), while HER2 status was not correlated (**Fig. 6C**). This is not the case for *CIZ1*-RD, arguing for a relationship between CIZ1-F and ERα and PR in primary tumours.

Discussion

Here we demonstrate that CIZ1-F is part of the RNA-dependent nuclear-matrix and elevated in early-stage cancers including hormone receptor negative, but not positive breast tumours. We show that CIZ1-F is distinct from other forms of CIZ1 and may have potential as a biomarker of early-stage disease.

CIZ1-F is part of the RNA-dependent NM and distinct from full-length CIZ1.

CIZ1 has multiple functions, both within and outside the context of DNA replication and the cell cycle, and its complexity is increased by the existence of multiple splice variants (**Supplementary Table 1**).[1-7, 18] We and others have shown that CIZ1 is a NM protein, defined by its resistance to extraction of chromatin,[3, 8] which suggests that its function is connected to nuclear architecture, possibly contributing to the spatial regulation of nuclear processes.[6-8, 18, 24] Although the NM appears to be corrupted in certain types of cancer[25] previous work[26-28] and our own experiments indicate that MCF-7 cells retain a NM structure to some extent. In MCF-7 cells both CIZ1-RD and CIZ1-F, as well as ERα, resist co-extraction with chromatin, indicating the presence of a non-chromatin stabilizing structure with which they are associated. This is of interest because CIZ1-F lacks the previously characterized NM anchor domain encoded by the C-terminal end of CIZ1. Moreover CIZ1-F is actually revealed by removal of chromatin, suggesting that the ARF-encoded epitope is normally in close proximity to chromatin, and masked by it. Further distinguishing CIZ1-F

from full-length CIZ1[8] is its behaviour in quiescent cells. Here, CIZ1-F is excluded from the nucleus and not evident even after removal of chromatin, suggesting that its role is exclusively in cycling cells. Finally, though CIZ1-F is co-extracted with the RNA fraction of the NM (**Fig. 2G**), it does not normally reside at the Xi implying that, unlike full-length CIZ1 this is not mediated by interaction with Xist lncRNA.[9, 10] Future research should now focus on assigning a function to CIZ1-F. Since the unique junction sequence is a very limited target, we were not able to achieve specific CIZ1-F knockdown using an siRNA approach (data not shown).

Biomarker potential for early stage cancers

Previous reports have shown overexpression of CIZ1 in colon,[29, 30] gallbladder,[21] lung,[31] and prostate cancer,[32] as well as ependymomas, gliomas, and medulloblastomas,[33] and it has also been reported that *CIZ1* knockout mice develop leukaemias and lymphomas.[9, 34] Moreover, splice variants of CIZ1 appear to have lineage-specific functions[4] some of which are over-represented in human tumours; notably CIZ1-Δ4 in Ewings tumour and CIZ1-B in lung tumours.[2, 5] However, for most published analyses it is not clear which variants are reported on, and in some cases transcript detection tools that are reported to reflect CIZ1 levels may actually reflect shifts in variant expression. Here we use validated and specific tools to demonstrate that *CIZ1*-F is upregulated in tumours of the colon and breast, that upregulation is most pronounced in early stage (I and II) disease, and is not directly correlated with proliferation. We also note that the association of *CIZ1*-F with tumour grade varies between tumour types, so although disrupted nuclear architecture and altered NM is a common feature of poorly differentiated and aggressive cancers,[25, 35] we cannot at this stage draw a correlation with suppression of CIZ1-F. In addition, our analyses have currently been limited to *CIZ1*-F mRNA levels. Future studies should determine CIZ1-F protein expression patterns in primary

tumour samples. Perhaps most useful is the clear correlation with hormone receptor status of breast tumours. Previous analysis has identified *CIZ1* as an oestrogen-responsive gene with oestrogen-response elements in its promoter.[12] Moreover, the N-terminal domain of CIZ1 protein can also interact with ER, conferring hypersensitivity to oestrogen in animal models and enhancing the tumourigenicity of breast cancer cells.[12] Though alternative splicing was not addressed in this study, a contribution of CIZ1-F to the cells response to oestrogen is likely because the ER interaction domain is partially retained in CIZ1-F, and it has two additional LXXLL nuclear receptor binding motifs[22] encoded by its unique C-terminal ARF. Contrary to published results[12] we did not observe induction of *CIZ1* upon stimulation with oestrogen for 24 h. A possible explanation for the apparent discrepancy is that different *CIZ1* primers were used previously,[12] which may well have reported on alternative splicing of *CIZ1* exon 8 rather than overall levels. In our study we detected *CIZ1*-RD and *CIZ1*-AD amplicons in regions unrelated to exon 8 in order to avoid this highly variable region of CIZ1. CIZ1 binding to (PR) has not been previously studied; though it has been shown that, in the context of NCoA-1/SRC-1, PR also requires two LXXLL domains.[36] Thus, while we could find no functional response to oestrogen in MCF-7 cells, the presence of these domains and the correlation with breast tumour receptor status supports a relationship *in vivo*.

Different CIZ1 splice variants at different cell proliferation stages

Transcript expression profile also distinguishes CIZ1-F from other variants of CIZ1. *CIZ1*-F has a distinct cell cycle-specific profile with low levels in G0, and delayed upregulation beyond the first cell cycle, which is at variance with the relatively constant expression of *CIZ1*-RD and *CIZ1*-AD amplicons (**Fig. 4E**). In these studies we also noted proliferation state dependent changes in the *CIZ1*-S and *CIZ1*-Δ4 variants. In fact in MCF-7 cells, they seem to be inversely related to *CIZ1*-F, with *CIZ1*-S and *CIZ1*-Δ4 being more highly

expressed in arrested cells (**Supplementary Fig. 3**). It is worth noting that both *CIZ1*-F and *CIZ1*-S-variants are the product of exon 8 splicing events, and that exon 8 splicing patterns have been implicated in Alzheimer's disease and cervical dystonia.[3, 14]

In conclusion, this exploratory study has convincingly distinguished CIZ1-F from CIZ1, revealed association with the RNA-dependent NM, and highlighted an unusual cell cycle expression pattern. It also supports a relationship with oestrogen responsive cancers and suggests that CIZ1-F may have potential as a biomarker.

Materials & Methods

Cell culture and tumour samples

MCF-7 breast cancer cells were cultured in minimal essential medium (cat. 21090-022; Gibco, Life Technologies, Thermo Fisher Scientific [TFS] Inc) in the presence of 10% foetal calf serum (FCS; cat. FB-5815 Biosera, Boussens, France) and 1X penicillin-streptomycin-glutamine (PSG; cat. 10378-016; Gibco, Life Technologies, TFS Inc). MRC5 normal lung foetal fibroblast cells were grown in Dulbecco's Modified Eagle Medium (cat. 21885-025; Gibco, Life Technologies, TFS Inc) in the presence of 10% FCS and 1X PSG. MCF-10A normal breast cells were cultured as previously described.[37] The cell lines were tested and authenticated in August 2017 using PCR-single-locus-technology (Eurofins Medigenomix Forensik GmbH). TissueScan^{TM} Cancer Survey cDNA array (CSRT103; Origene) was used to assess *CIZ1*-F levels in different human cancer types. The breast cancer cDNA array BCRT102 (Origene) was used to screen 43 breast cancer samples and 5 normal breast tissues. The colon cancer cDNA array HCRT103 (Origene) was used to screen 24 colon cancer samples with matched normal tissues.

Quantitative real-time PCR (qRT-PCR)

RNA was isolated using Trizol® Reagent (cat. 15596-026; Ambion RNA, Life Technologies, TFS Inc) and reverse transcribed into cDNA using Superscript® III First-Strand Synthesis System (cat. 18080-051; Life Technologies, TFS Inc). Quantitative RT-PCR was performed using SYBR green (FastSYBR® Green Mastermix, cat. 4385612; Applied Biosystems® [AB], TFS Inc) or Taqman (Taqman® Fast Universal PCR Master Mix, cat. 4352042; AB, TFS Inc) reagents using the AB StepOnePlus™ Real-Time PCR Systems (TFS Inc). The following program was used: 95°C for 20 s, followed by 40 cycles of 95°C for 3 s and 60°C for 30 s. Data was analysed using the StepOne Software v. 2.3 (AB, TFS Inc). Origene® cancer cDNA array plates were processed with the AB 7300 system (TFS Inc) using the TaqMan® Universal Master Mix II (cat. 4440040; AB, TFS Inc) using the following program: 50°C for 2 min, 95°C for 10 min, followed by 50 cycles of 95°C for 15 s and 60°C for 1 min.

Primers were designed using Primer 3 Plus (http://primer3plus.com/), except for the *ACTB* primers[5] and *CYPA* primers,[38] which were previously described, and cyclin D1 and cyclin E1 primers, which were retrieved from qPrimerDepot (http://primerdepot.nci.nhi.gov). All primers are listed in **Supplementary Table 2**. *CIZ1*-RD was detected with primers in exon 6/7 and *CIZ1*-AD with primers in exon 15/16 (**Fig. 1A, Supplementary Table 2**). For the Origene cDNA array plates, *CIZ1*-F was detected with primers 23-25, *CIZ1*-RD with primers 20-22 and *MKI67* using a Taqman Assay (assay no. Hs01032443_m1). All expression levels have been normalised to the mean expression of *ACTB* and *CYPA* unless stated differently.

Constructs and transfection

Human *CIZ1*-F was cloned from cDNA from MCF-7 cells using PCR with DreamTaq (cat. EP1701; TFS Inc) and T/A cloning into the pGEM®-T Easy Vector System I (cat. A1360; Promega) and verified by sequencing. *CIZ1*-F was subsequently transferred using restriction enzymes (PmlI and SanDI, cat. ER0361 and FD2164, respectively; TFS Inc) into the pEGFP-C3 vector, allowing direct comparison with full-length *CIZ1*, containing all translated exons (2-17), also cloned into pEGFP-C3 as part of a previous study.[5] Both plasmids were transiently transfected into MCF-7 cells using the *Trans*IT®-3T3 and *Trans*IT®-LT1 Transfection Kits (cat. MIR 2180 and MIR 2300, respectively; Mirus Bio LLC).

Antibodies

CIZ1-F specific antibody was made by immunising rabbits with two peptides (**Fig. 1C**) from the unique CIZ1-F ARF-encoded peptide (Covalab UK Ltd). Antisera from two rabbits were pooled and immunopurified against peptide 2 (which had no significant similarity to any other protein) and used for Western Blot and immunofluorescence (dilutions 1:100-1:500 and 1:50, respectively). CIZ1-F blocking peptides were used in a 1:500 dilution. Mouse monoclonal antibody (hC221a) against CIZ1 was raised by Fusion Antibodies against a recombinant human CIZ1 C-terminal fragment encoded by amino acids 678-898 in reference sequence UniProt Q9ULV3. In addition, purified CIZ1 1794 polyclonal antiserum[6] was used to detect CIZ1 without discriminating CIZ1-F. For immunofluorescence goat polyclonal anti-mouse-AlexaFluor488 conjugate or goat polyclonal anti-rabbit-AlexaFluor568 conjugate (cat. A11001 and A11011, respectively, Life Technologies, TFS) were used as appropriate. For Western Blot CIZ1 exon 5 antibody (cat. HPA-020380; Sigma-Aldrich) was used to detect CIZ1 without discriminating CIZ1-F. Actin (cat. A3853, Sigma Aldrich) was used as a loading control.

Immunodetection

Protein extracts were obtained by washing cells with cold PBS and harvesting directly into SDS sample buffer (100 mM DTT, 2% SDS, 60 mM Tris pH 6.8, 0.001% bromophenol blue) at 95°C in the presence of 2 mM PMSF. Western blot was performed as previously described.[6] Bands were visualised using EZ-ECL (cat. 20-500-120; Biological Industries Ltd).

For immunofluorescence, cells grown on glass coverslips were transferred into PBS, washed with cytoskeletal (CSK) buffer (containing EDTA-free protease inhibitor tablets [cat. 05056489001; F Hoffmann-La Roche Ltd, Sigma Aldrich] and 1 mM DTT) with or without 0.1% Triton X-100[39] for 30 s, then transferred to PBS and fixed for 20 min in 4% paraformaldehyde in PBS. Primary antibodies were incubated for 1.5 h at 37°C and secondary antibodies for 1 h at 37°C in the dark in antibody buffer (PBS, 10 mg/ml BSA, 0.2% SDS, 0.1% Triton X-100). Nuclei were counterstained with DAPI (VECTASHIELD Antifade Mounting Medium with DAPI, cat. H-1200; Vector Laboratories).

Fluorescent image data was collected with a Zeiss Axiovert microscope (Carl Zeiss) with a 63/1.40 oil immersion objective and an AxioCam camera (Carl Zeiss Vision) with Openlab software (version 4.0.2, Improvision). Images were processed using Adobe PhotoShop to reduce background to black and to overlay multicolor images. Where quantitative data is derived from the images, all capture parameters and image processing was identical for samples and controls within an experiment.

Nuclear Matrix extractions

Cells were fractionated as previously described.[39, 40] In brief, coverslips were washed with detergent (0.1% Triton-X-100 in CSK), followed by the same buffer with 0.5 M NaCl. Cells were then incubated at 37°C for 1 h in DNase Buffer with or without DNase I (0.5 U/µl, cat.

000000004716728001; F Hoffmann-La Roche Ltd, Sigma Aldrich), RNase (cat. 000000011579681001; F Hoffmann-La Roche Ltd, Sigma Aldrich) or both nucleases in combination, then fixed for 20 min with 8% paraformaldehyde. Vanadyl Ribonucleoside Complex (cat. S1402S; New England Biolabs) was added at 1:80 to all buffers with the exception of the RNase containing treatment step and its preceding wash steps. Coverslips were counterstained with Hoechst33258 (1:100,000, cat. H3569; TFS) to verify removal of chromatin, and immunostained as indicated.

Cell cycle synchronisation

MCF-7 cells were synchronised by release from quiescence as described [41]. Cells were grown to confluence, then the medium was changed, and cells cultured for a further four days without further media changes and then released in serum-containing medium by 1:4 splitting. Alternatively, cells were incubated in the presence of nocodazole (0.04 µg/ml), aphidicolin (5 µg/ml) or thymidine (2.5 mM) for 24 h, and released from arrest by washing twice in PBS followed by addition of fresh medium. In order to verify synchrony, incorporation of EdU into newly synthesized DNA was measured using Click-iT® EdU Alexa Fluor® 555 (cat. C10338; Invitrogen, Life Technologies, TFS Inc).

Statistical analysis

Statistical analysis was performed using IBM SPSS Statistics for Windows (version 21.0.0.0). Associations between expression levels and clinical parameters were calculated using the Mann Whitney U-test or the Kruskall-Wallis test, and for related samples using the Related-Samples Wilcoxon Signed Rank Test, as indicated. Relationships between expression levels amongst different genes were determined using Pearson Correlation and linear regression. All statistical tests were two-sided and all *P*-values were considered

significant when < 0.05. Significance levels were indicated as follows: *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$.

List of abbreviations

AB, Applied Biosystems; AD, anchor domain; ARF, alternative reading frame; CIZ1, CDKN1A-interacting zinc finger protein; CIZ1-B, CIZ1 b-variant; CIZ1-Δ4, CIZ1 Δ4-variant; CIZ1-F, CIZ1 f-variant; CIZ1-M, CIZ1 m-variant; CIZ1-S, CIZ1 s-variant; CSK, cytoskeletal; ERα, oestrogen receptor alpha; FCS, foetal calf serum; NM, nuclear matrix; PR, progesterone receptor; PSG, penicillin-streptomycin-glutamine; RD, replication domain; RQ, relative quantity; SD, standard deviation; SEM, standard error of mean; qRT-PCR, quantitative real-time PCR; TFS, Thermo Fisher Scientific; Xi, inactive X chromosome; Xist lncRNA, X-inactive specific transcript long non-coding RNA.

Author's contribution

DS and ES conducted experiments. GH provided technical assistance. DS and DC wrote the manuscript.

Acknowledgements

We thank Dr Justin Ainscough for critical comments on the manuscript, Hannah Screeton for her contribution to cell studies and quantitative PCRs, and Dr William Brackenbury for providing cell line RNA.

Figure Legends

Figure 1. Human CIZ1 and variant CIZ1-F.

(A) Exonic sequence in messenger RNA of full-length and *CIZ1*-F, with location of primers (pri) and Taqman probes indicated. Sequences are provided in **Supplementary Table 2**. Lines indicate the location of sequences that encode the protein domains illustrated in (B). Four alternative exon 1s are indicated; *CIZ1*-F has been detected in combination with exons 1b, 1c, and 1d.

(B) Full-length and CIZ1-F protein showing annotated protein domains including glutamine-rich regions (Q), nuclear localisation signals (NLS), zinc finger domains (ZF), matrin-3 domain (MH3), and acidic domain (AcD). Characterised functional domains and interaction sites for full-length CIZ1 are also shown, including replication domain (RD) which contains all sequences required for replication activity,[6] and anchor domain (AD) which contains sequences that are sufficient to mediate attachment to the nuclear matrix.[8] The CIZ1-F specific sequence, encoded by an alternative reading frame (ARF) of exon 12-13, is highlighted in light blue. Also shown are the location of K/RXL cyclin-binding motifs implicated in replication[7] ('C'), LXXLL ER-interaction motifs (asterisks), and sequences reported for full-length CIZ1 to interact with CDC6,[18] cyclin-dependent kinase 2 (CDK2),[15] cyclins A and E,[7] dynein light chain (DYNLL1),[15] oestrogen-receptor (ER),[12] enhancer of rudimentary homolog (ERH),[42] p21,[16] and YAP.[17] Note that several of these interactions were discovered in mouse CIZ1, and that the reported interaction between ER and CIZ1 is not via a domain that includes LXXLL.[12] X-axis shows amino acid number.

(C) Amino acid sequence of the CIZ1-F protein, with the sequence specified by its ARF bold and underlined. The peptides in red and purple were used to generate CIZ1-F-specific antibodies.

Figure 2. Nuclear CIZ1-F resists extraction of chromatin but not RNA.

(A-B) CIZ1-F does not localise to the inactive X chromosome (Xi). Immunofluorescence with a monoclonal antibody for a peptide located in the CIZ1 anchor domain (AD; green) shows both diffuse nuclear expression and specific localisation to a region previously shown to be the Xi.[9] Immunofluorescence with purified antibody raised to the CIZ1-F ARF (red) shows absence of Xi-localisation. DNA is stained with Hoechst33258 (blue). Nuclei in white boxes are enlarged in (B).

(C) Overview of nuclear matrix (NM) extraction procedure. MCF-7 cells were serially extracted with 1) detergent-containing cytoskeletal buffer (CSK), 2) detergent-containing CSK supplemented with 0.5 M NaCl (mock extraction), 3) DNase 1, or 4) RNase, or 5) both enzymes (see methods).

(D) MCF-7 cells subjected to the treatments described under (C) were counterstained with Hoechst33258 to control for removal of chromatin (blue), and probed with purified polyclonal anti-CIZ1-F or replication domain (CIZ1-RD) antibody 1794 (red). CIZ1-RD resists all treatments, while CIZ1-F is sensitive to extraction of RNA, and is also dramatically revealed by removal of chromatin. Bar is 10 microns.

(E) Quantification of the fluorescence intensities in (D), shown after subtraction of background signal and expressed relative to detergent treated cells. Number of cells quantified $\geq$ 70 per condition. Numbers refer to the treatment conditions listed above.

(F) Enhanced and enlarged images of the indicated nuclei i-iv from D, showing detergent and DNase-resistant nuclear fractions.

(G) Interpretation of the data showing the dependency of CIZ1-F on RNA for nuclear retention, compared to full resistance of CIZ1-RD.

Figure 3. CIZ1-F transcript is suppressed and CIZ1-F protein is excluded from the nuclear matrix in quiescent cells.

(A) Immunodetection of CIZ1-F in cycling (day 2), confluent (day 4) and quiescent (day 7 with medium change) populations of MCF-7 cells, after washing with detergent-cytoskeletal buffer to remove soluble protein. Enlarged images of example cells are shown and DAPI-stained images (left) show cell nuclei in the same fields. Bar is 10 microns.

(B) Quantification of the average immunofluorescence signal of cells depicted in (A) after subtraction of background. Abbreviations: Cycl, cycling; Confl, confluent; Qsct, quiescent.

(C) NM-extraction of quiescent MCF-7 cell populations showing no CIZ1-F on the NM after treatment with DNase. For details on the method see **Fig. 2A**.

(D) Transcript levels in MCF-7 breast epithelial carcinoma cells at the indicated number of days post-plating, measured by quantitative RT-PCR (qPCR). A parallel culture that received regular changes of media (indicated as 'MC'), harvested at 11 days, is shown for comparison. Histograms show the mean of three technical replicates ± SEM of a representative experiment (experiments were repeated at least 3 times). Data is expressed as relative quantification (RQ) after normalization to *ACTB* and *CYPA*, and is calibrated to levels at day 2. Cell counts at the time of harvesting are provided ± SEM (black dotted line; average of three counts expressed as proportion of a confluent population). Right, comparison between replication domain (CIZ1-RD), anchor domain (CIZ1-AD) and CIZ1-F in MCF-7 cells at day 2 ± SEM, with no normalization to show relative levels.

(E) As in (D) for MRC5 normal foetal lung fibroblast cells. Right, comparison between *CIZ1*-RD, *CIZ1*-AD and *CIZ1*-F at day 2. SEM for three technical replicates is shown.

Primer and probe sequences are in **Supplementary Table 2**.

Figure 4. *CIZ1*-F transcript is expressed in G1 phase.

(A) Quantitative RT-PCR showing *CIZ1*-F transcript (red) in MCF-7 cells during the first cell cycle following release from contact-inhibition and serum-deprivation induced cell cycle arrest, relative to *ACTB* and *CYPA* housekeeping genes, and calibrated to unreleased cells. *CIZ1* anchor domain (*CIZ1*-AD) and replication domain (*CIZ1*-RD) expression (grey lines), as well as cyclin E1 (blue) are shown for comparison. Data show means of four experiments ± SD. The percentage of cells that incorporated EdU into newly synthesized DNA during a 30 min pulse at the indicated times is shown for comparison (black line; two biological repeats ± SD). Note that error bars for cyclin E1 are very small (range, 0.00001-0.008). Cell cycle stages (estimates based on EdU-incorporation and cyclin-expression) are indicated below. Images on the right show EdU (purple) and total DNA (blue) before (0 h) and after (20 h) release. Bar is 10 microns.

(B) *CIZ1*-AD, *CIZ1*-RD and *CIZ1*-F transcript in MCF-7 cells 31 h post-release from quiescence, in the presence and absence of aphidicolin (Aph), nocodazole (Noc) and thymidine (Thy), showing mean of three biological replicates ± SEM.

(C) As in (B), but for MRC5 normal lung foetal fibroblast cells. Mean RQ of three technical replicates is expressed relative to cycling samples (± SEM).

(D) As in (A), but showing release of MCF-7 cells from a 24 h nocodazole-arrest applied 16 h after release from quiescence. Synchronisation strategy is indicated above the graph. Graph shows transcript levels (mean of three technical replicates ± SEM), at the indicated hours after release from arrest, with cell cycle stage (based on cyclin E expression) illustrated below. Right, control sample without nocodazole (40 h post-release from quiescence).

(E) Schematic summarising *CIZ1*-RD, *CIZ1*-AD and *CIZ1*-F mRNA expression levels during the first cell cycle following release from cell cycle arrest (Q-G1) and mitotic arrest (M-G1). Primer and probe sequences are in **Supplementary Table 2**.

Figure 5. *CIZ1*-F is overexpressed in early stage human breast and colon cancer.

(A) *CIZ1*-F and replication domain (*CIZ1*-RD) expression in 24 primary colon tumours and matched normal samples (Origene colon cancer cDNA array HCRT103), showing mean mRNA expression levels for the average of all matched normal tissues and the four cancer stage classifications. *CIZ1*-RD and *MKI67* are shown for comparison. Individual data per case can be found in **Supplementary Fig. 4**, and individual classifications and pathology notes accesses at www.origene.com.

(B) As in (A), showing 5 normal samples and 43 primary breast cancer samples from the indicated stages (Origene breast cancer cDNA array BCRT102). Data for individual cases can be found in **Supplementary Fig. 5**.

(C) As in (A) for the same 24 colon cancer samples and matched normal samples, showing mean mRNA expression grouped by grade. Grade 1: 50% late stage; grade 2: 38% late stage; grade 3: 100% late stage (1 case, white bar). Late stage refers to stage III and stage IV.

(D) As in (B) for the same 5 normal samples and 43 primary breast cancer samples, showing mean mRNA expression grouped by grade. Grade 1: 100% late stage tumours; grade 2: 43% late stage tumours; grade 3: 27% late stage tumours. Late stage refers to stage III and stage IV.

Mean RQ's to the mean *CIZ1*-RD-expression of normal samples ± SEM are shown. Significant differences are indicated (NS, not significant). Primer and probe sequences are in **Supplementary Table 2**.

Figure 6. *CIZ1*-F expression is increased in ER-negative tumours.

Box-plots showing *CIZ1* replication domain (*CIZ1*-RD; left), *CIZ1*-F (middle) and *MKI67* (right) expression in (A) normal samples and oestrogen receptor (ER)-positive and -negative

tumours, (B) normal samples and progesterone receptor (PR)-positive and -negative tumours, and (C) normal samples and HER-2-positive and -negative tumours. RQ's are expressed relative to mean *CIZ1*-RD-expression of normal samples. Significant differences between subgroups are indicated (Mann Whitney U-tests; NS, not significant). Primer and probe sequences are in **Supplementary Table 2**.

Reference list

1. Rahman FA, Aziz N, Coverley D. Differential detection of alternatively spliced variants of Ciz1 in normal and cancer cells using a custom exon-junction microarray. BMC Cancer 2010; 10:482.

2. Rahman F, Ainscough JF, Copeland N, Coverley D. Cancer-associated missplicing of exon 4 influences the subnuclear distribution of the DNA replication factor CIZ1. Hum Mutat 2007; 28:993-1004.

3. Dahmcke CM, Buchmann-Moller S, Jensen NA, Mitchelmore C. Altered splicing in exon 8 of the DNA replication factor CIZ1 affects subnuclear distribution and is associated with Alzheimer's disease. Mol Cell Neurosci 2008; 38:589-94.

4. Greaves EA, Copeland NA, Coverley D, Ainscough JF. Cancer-associated variant expression and interaction of CIZ1 with cyclin A1 in differentiating male germ cells. Journal of cell science 2012; 125:2466-77.

5. Higgins G, Roper KM, Watson IJ, Blackhall FH, Rom WN, Pass HI, Ainscough JF, Coverley D. Variant Ciz1 is a circulating biomarker for early-stage lung cancer. Proc Natl Acad Sci U S A 2012; 109:E3128-35.

6. Coverley D, Marr J, Ainscough J. Ciz1 promotes mammalian DNA replication. Journal of cell science 2005; 118:101-12.

7. Copeland NA, Sercombe HE, Ainscough JF, Coverley D. Ciz1 cooperates with cyclin-A-CDK2 to activate mammalian DNA replication in vitro. Journal of cell science 2010; 123:1108-15.

8. Ainscough JF, Rahman FA, Sercombe H, Sedo A, Gerlach B, Coverley D. C-terminal domains deliver the DNA replication factor Ciz1 to the nuclear matrix. Journal of cell science 2007; 120:115-24.

9. Ridings-Figueroa R, Stewart ER, Nesterova TB, Coker H, Pintacuda G, Godwin J, Wilson R, Haslam A, Lilley F, Ruigrok R, et al. The nuclear matrix protein CIZ1 facilitates localization of Xist RNA to the inactive X-chromosome territory. Genes & development 2017; 31:876-88.

10. Sunwoo H, Colognori D, Froberg JE, Jeon Y, Lee JT. Repeat E anchors Xist RNA to the inactive X chromosomal compartment through CDKN1A-interacting protein (CIZ1). Proc Natl Acad Sci U S A 2017; 114:10654-9.

11. Coverley D, Higgins G, West D, Jackson OT, Dowle A, Haslam A, Ainscough E, Chalkley R, White J. A quantitative immunoassay for lung cancer biomarker CIZ1b in patient plasma. Clin Biochem 2017; 50:336-43.

12. den Hollander P, Rayala SK, Coverley D, Kumar R. Ciz1, a Novel DNA-binding coactivator of the estrogen receptor alpha, confers hypersensitivity to estrogen action. Cancer research 2006; 66:11021-9.

13. Siegel R, Ma J, Zou Z, Jemal A. Cancer statistics, 2014. CA: a cancer journal for clinicians 2014; 64:9-29.

14. Xiao J, Uitti RJ, Zhao Y, Vemula SR, Perlmutter JS, Wszolek ZK, Maraganore DM, Auburger G, Leube B, Lehnhoff K, et al. Mutations in CIZ1 cause adult onset primary cervical dystonia. Ann Neurol 2012; 71:458-69.

15. den Hollander P, Kumar R. Dynein light chain 1 contributes to cell cycle progression by increasing cyclin-dependent kinase 2 activity in estrogen-stimulated cells. Cancer research 2006; 66:5941-9.

16. Mitsui K, Matsumoto A, Ohtsuka S, Ohtsubo M, Yoshimura A. Cloning and characterization of a novel p21(Cip1/Waf1)-interacting zinc finger protein, ciz1. Biochem Biophys Res Commun 1999; 264:457-64.

17. Lei L, Wu J, Gu D, Liu H, Wang S. CIZ1 interacts with YAP and activates its transcriptional activity in hepatocellular carcinoma cells. Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine 2016; 37:11073-9.

18. Copeland NA, Sercombe HE, Wilson RH, Coverley D. Cyclin A/CDK2 phosphorylation of CIZ1 blocks replisome formation and initiation of mammalian DNA replication. Journal of cell science 2015; 128:1518-27.

19. Jiang L, Luo X, Shi J, Sun H, Sun Q, Sheikh MS, Huang Y. PDRG1, a novel tumor marker for multiple malignancies that is selectively regulated by genotoxic stress. Cancer Biol Ther 2011; 11:567-73.

20. Thalappilly S, Suliman M, Gayet O, Soubeyran P, Hermant A, Lecine P, Iovanna JL, Dusetti NJ. Identification of multi-SH3 domain-containing protein interactome in pancreatic cancer: a yeast two-hybrid approach. Proteomics 2008; 8:3071-81.

21. Zhang D, Wang Y, Dai Y, Wang J, Suo T, Pan H, Liu H, Shen S, Liu H. CIZ1 promoted the growth and migration of gallbladder cancer cells. Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine 2015; 36:2583-91.

22. Heery DM, Kalkhoven E, Hoare S, Parker MG. A signature motif in transcriptional co-activators mediates binding to nuclear receptors. Nature 1997; 387:733-6.

23. Chaligne R, Popova T, Mendoza-Parra MA, Saleem MA, Gentien D, Ban K, Piolot T, Leroy O, Mariani O, Gronemeyer H, et al. The inactive X chromosome is epigenetically unstable and transcriptionally labile in breast cancer. Genome research 2015; 25:488-503.

24. Wilson RH, Coverley D. Relationship between DNA replication and the nuclear matrix. Genes Cells 2013; 18:17-31.

25. Zink D, Fischer AH, Nickerson JA. Nuclear structure in cancer cells. Nat Rev Cancer 2004; 4:677-87.

26. Spencer VA, Samuel SK, Davie JR. Nuclear matrix proteins associated with DNA in situ in hormone-dependent and hormone-independent human breast cancer cell lines. Cancer research 2000; 60:288-92.

27. Samuel SK, Minish TM, Davie JR. Nuclear matrix proteins in well and poorly differentiated human breast cancer cell lines. J Cell Biochem 1997; 66:9-15.

28. Khanuja PS, Lehr JE, Soule HD, Gehani SK, Noto AC, Choudhury S, Chen R, Pienta KJ. Nuclear matrix proteins in normal and breast cancer cells. Cancer research 1993; 53:3394-8.

29. Wang DQ, Wang K, Yan DW, Liu J, Wang B, Li MX, Wang XW, Liu J, Peng ZH, Li GX, et al. Ciz1 is a novel predictor of survival in human colon cancer. Experimental biology and medicine 2014; 239:862-70.

30. Yin J, Wang C, Tang X, Sun H, Shao Q, Yang X, Qu X. CIZ1 regulates the proliferation, cycle distribution and colony formation of RKO human colorectal cancer cells. Molecular medicine reports 2013; 8:1630-4.

31. Zhou X, Liu Q, Wada Y, Liao L, Liu J. CDKN1A-interacting zinc finger protein 1 is a novel biomarker for lung squamous cell carcinoma. Oncology letters 2018; 15:183-8.

32. Liu T, Ren X, Li L, Yin L, Liang K, Yu H, Ren H, Zhou W, Jing H, Kong C. Ciz1 promotes tumorigenicity of prostate carcinoma cells. Front Biosci (Landmark Ed) 2015; 20:705-15.

33. Warder DE, Keherly MJ. Ciz1, Cip1 interacting zinc finger protein 1 binds the consensus DNA sequence ARYSR(0-2)YYAC. J Biomed Sci 2003; 10:406-17.

34. Nishibe R, Watanabe W, Ueda T, Yamasaki N, Koller R, Wolff L, Honda Z, Ohtsubo M, Honda H. CIZ1, a p21Cip1/Waf1-interacting protein, functions as a tumor suppressor in vivo. FEBS Lett 2013; 587:1529-35.

35. Verheijen R, van Venrooij W, Ramaekers F. The nuclear matrix: structure and composition. Journal of cell science 1988; 90 (Pt 1):11-36.

36. McInerney EM, Rose DW, Flynn SE, Westin S, Mullen TM, Krones A, Inostroza J, Torchia J, Nolte RT, Assa-Munt N, et al. Determinants of coactivator LXXLL motif specificity in nuclear receptor transcriptional activation. Genes & development 1998; 12:3357-68.

37. Tait L, Soule HD, Russo J. Ultrastructural and immunocytochemical characterization of an immortalized human breast epithelial cell line, MCF-10. Cancer research 1990; 50:6087-94.

38. Swarts DR, Henfling ME, Van Neste L, van Suylen RJ, Dingemans AM, Dinjens WN, Haesevoets A, Rudelius M, Thunnissen E, Volante M, et al. CD44 and OTP are strong prognostic markers for pulmonary carcinoids. Clinical cancer research : an official journal of the American Association for Cancer Research 2013; 19:2197-207.

39. Wilson RHH, E.L.;Coverley,D. The Nuclear Matrix: Fractionation Techniques and Analysis. In: Pryor PP, ed. Subcellular Fractionation - A Laboratory Manual. New York: Cold Spring Harbor Laboratory Press, 2015:223-33.

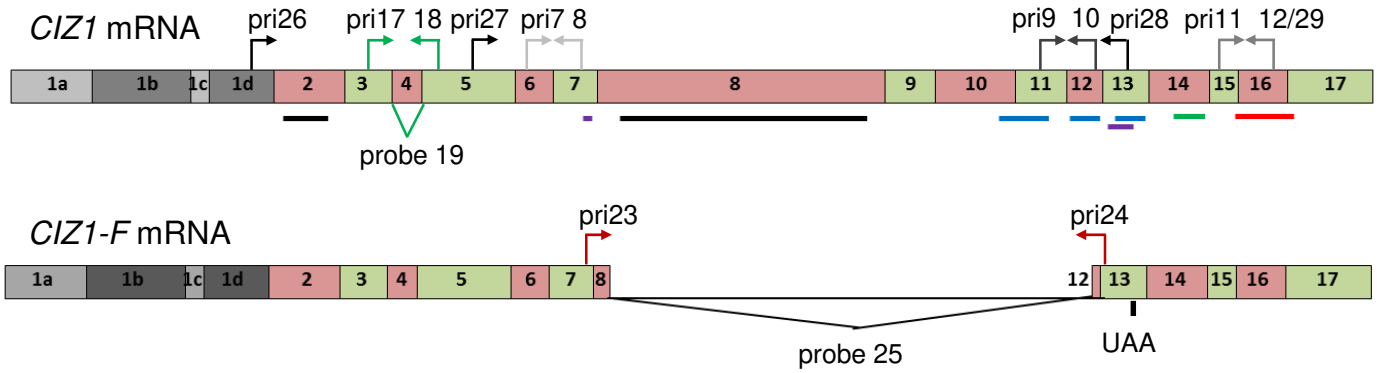
40. Hesketh EL, Knight JR, Wilson RH, Chong JP, Coverley D. Transient association of MCM complex proteins with the nuclear matrix during initiation of mammalian DNA replication. Cell Cycle 2015; 14:333-41.

41. Coverley D, Laman H, Laskey RA. Distinct roles for cyclins E and A during DNA replication complex assembly and activation. Nature cell biology 2002; 4:523-8.

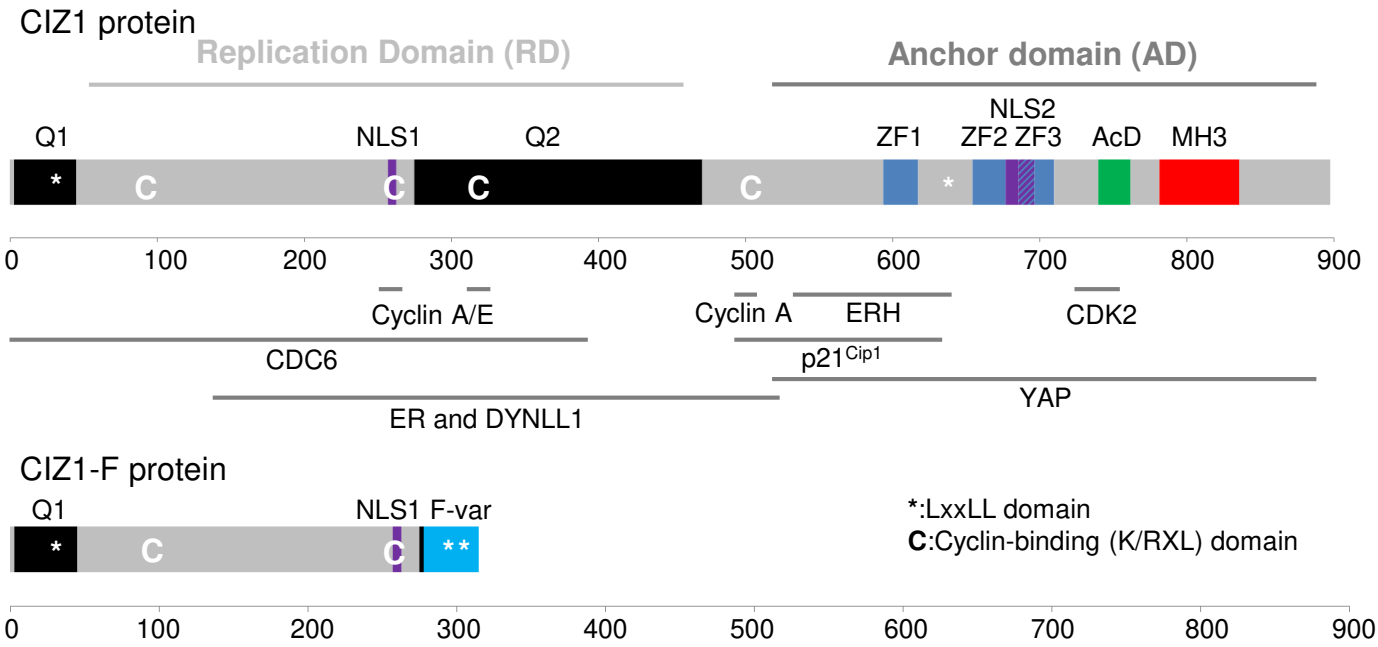
42. Lukasik A, Uniewicz KA, Kulis M, Kozlowski P. Ciz1, a p21 cip1/Waf1-interacting zinc finger protein and DNA replication factor, is a novel molecular partner for human enhancer of rudimentary homolog. FEBS J 2008; 275:332-40.

Figure 1

A



B



C

M F S Q Q Q Q Q Q L Q Q Q Q Q Q L Q Q L Q Q Q Q L Q Q Q Q L Q Q Q Q L L Q L Q Q L L Q Q S P P Q
A P L P M A V S R G L P P Q Q P Q Q P L L N L Q G T N S A S L L N G S M L Q R A L L L Q Q L Q G L D
Q F A M P P A T Y D T A G L T M P T A T L G N L R G Y G M A S P G L A A P S L T P P Q L A T P N L Q
Q F F P Q A T R Q S L L G P P P V G V P M N P S Q F N L S G R N P Q K Q A R T S S S T T P N R K D S
S S Q T M P V E D K S D P P E G S E E A A E P R M D T P E D Q D L P P C P E D I A K E K R T P A P E
P E P C E A S E L P A K R L R S S E E P T E K E P P G Q L Q **D T G P Q D C Q T I L A T L L H R L Q P L**
L Q N P S Q V C G A R E V P G A

Figure 2

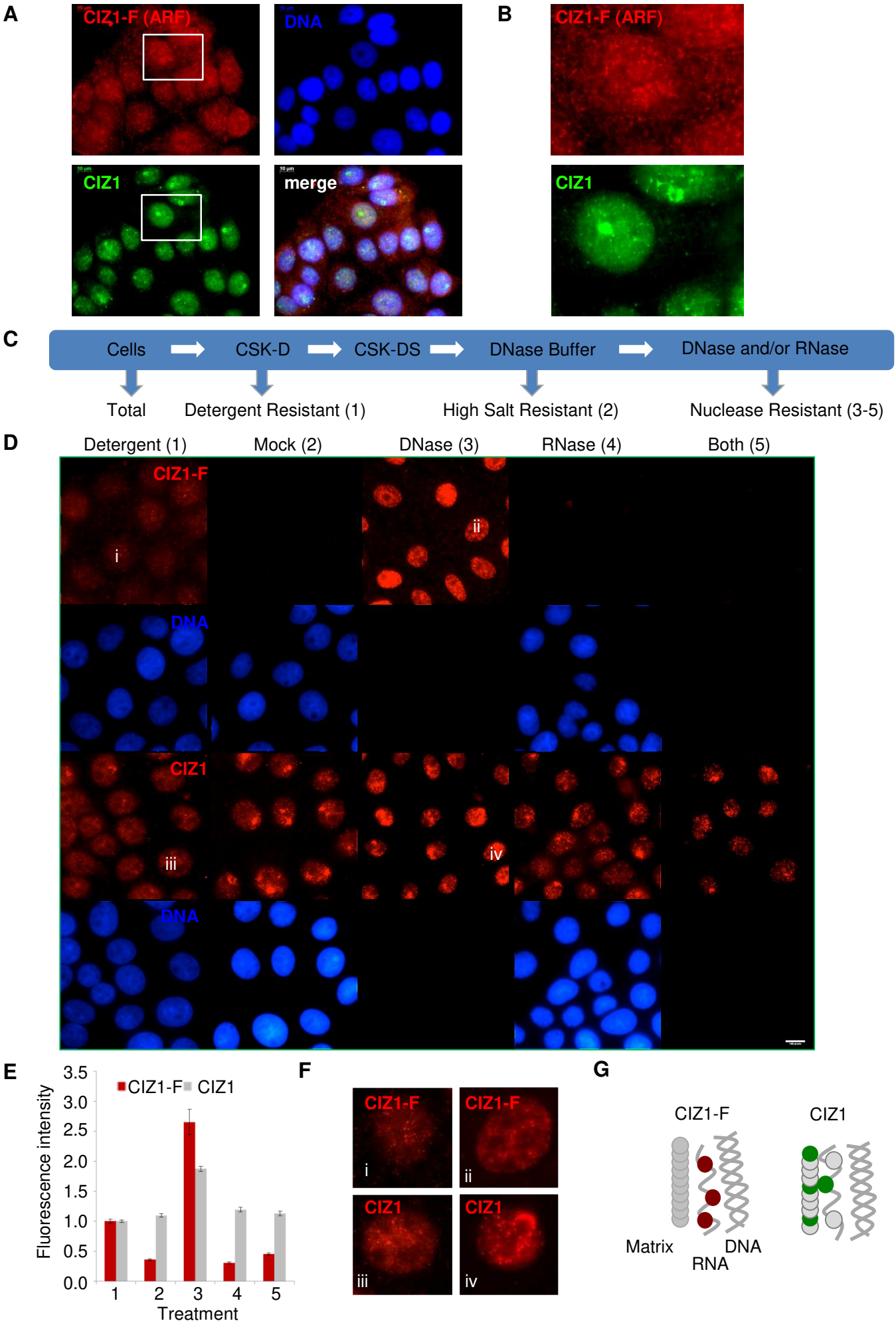

A
CIZ1-F (ARF)
DNA
CIZ1
merge
B
CIZ1-F (ARF)
CIZ1
C
Cells
CSK-D
CSK-DS
DNase Buffer
DNase and/or RNase
Total
Detergent Resistant (1)
High Salt Resistant (2)
Nuclease Resistant (3-5)
D
Detergent (1)
Mock (2)
DNase (3)
RNase (4)
Both (5)
CIZ1-F
DNA
CIZ1
DNA
i
ii
iii
iv
E
CIZ1-F
CIZ1
Fluorescence intensity
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
1
2
3
4
5
Treatment
F
CIZ1-F
CIZ1-F
CIZ1
CIZ1
i
ii
iii
iv
G
CIZ1-F
CIZ1
Matrix
DNA
RNA

Figure 3

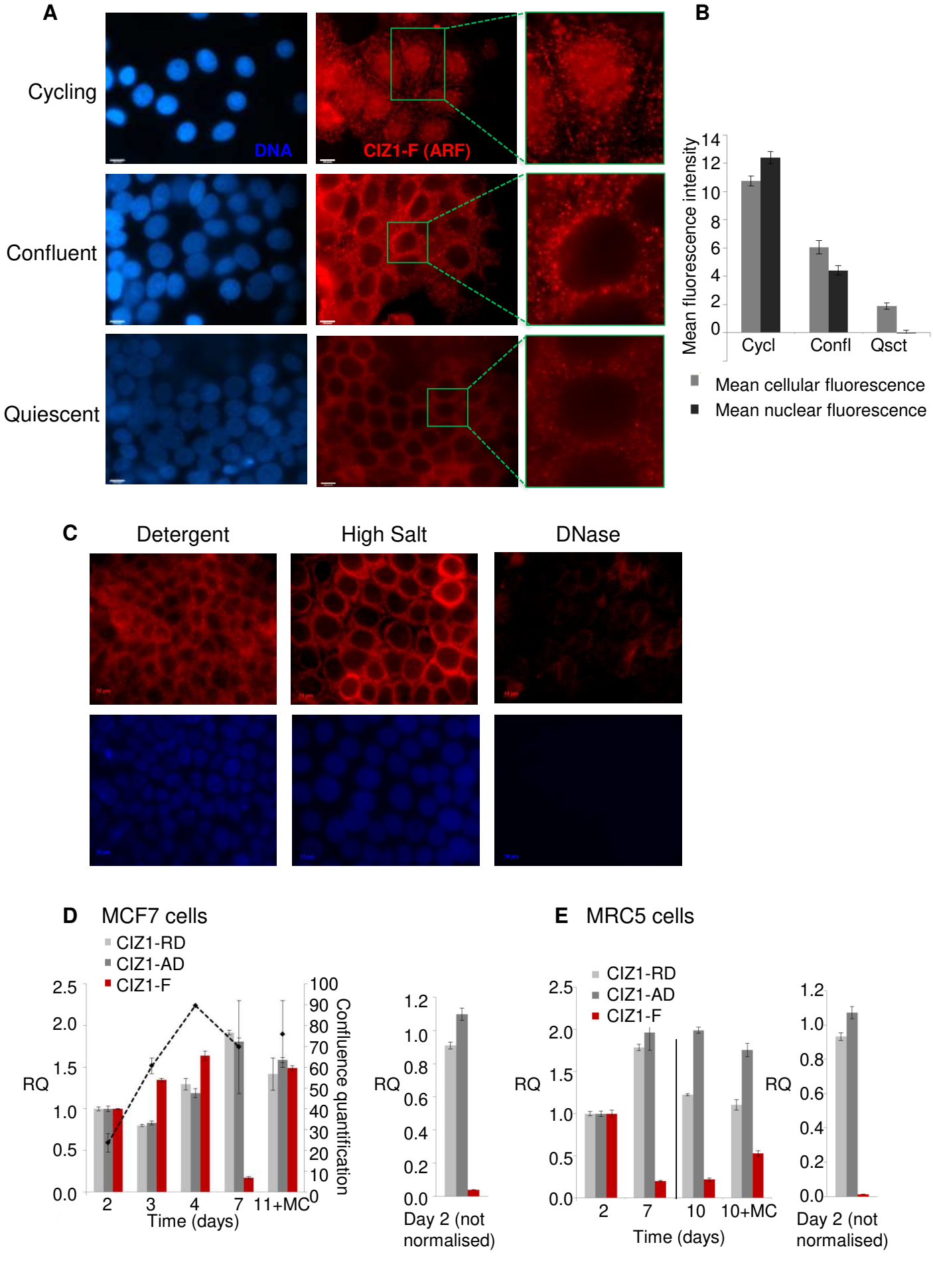

A
Cycling
Confluent
Quiescent
DNA
CIZ1-F (ARF)
B
Mean fluorescence intensity
14
12
10
8
6
4
2
0
Cycl
Confl
Qsct
Mean cellular fluorescence
Mean nuclear fluorescence
C
Detergent
High Salt
DNase
D
MCF7 cells
CIZ1-RD
CIZ1-AD
CIZ1-F
RQ
Confluence quantification
Time (days)
2
3
4
7
11+MC
Day 2 (not normalised)
E
MRC5 cells
CIZ1-RD
CIZ1-AD
CIZ1-F
RQ
Time (days)
2
7
10
10+MC
Day 2 (not normalised)

Figure 4

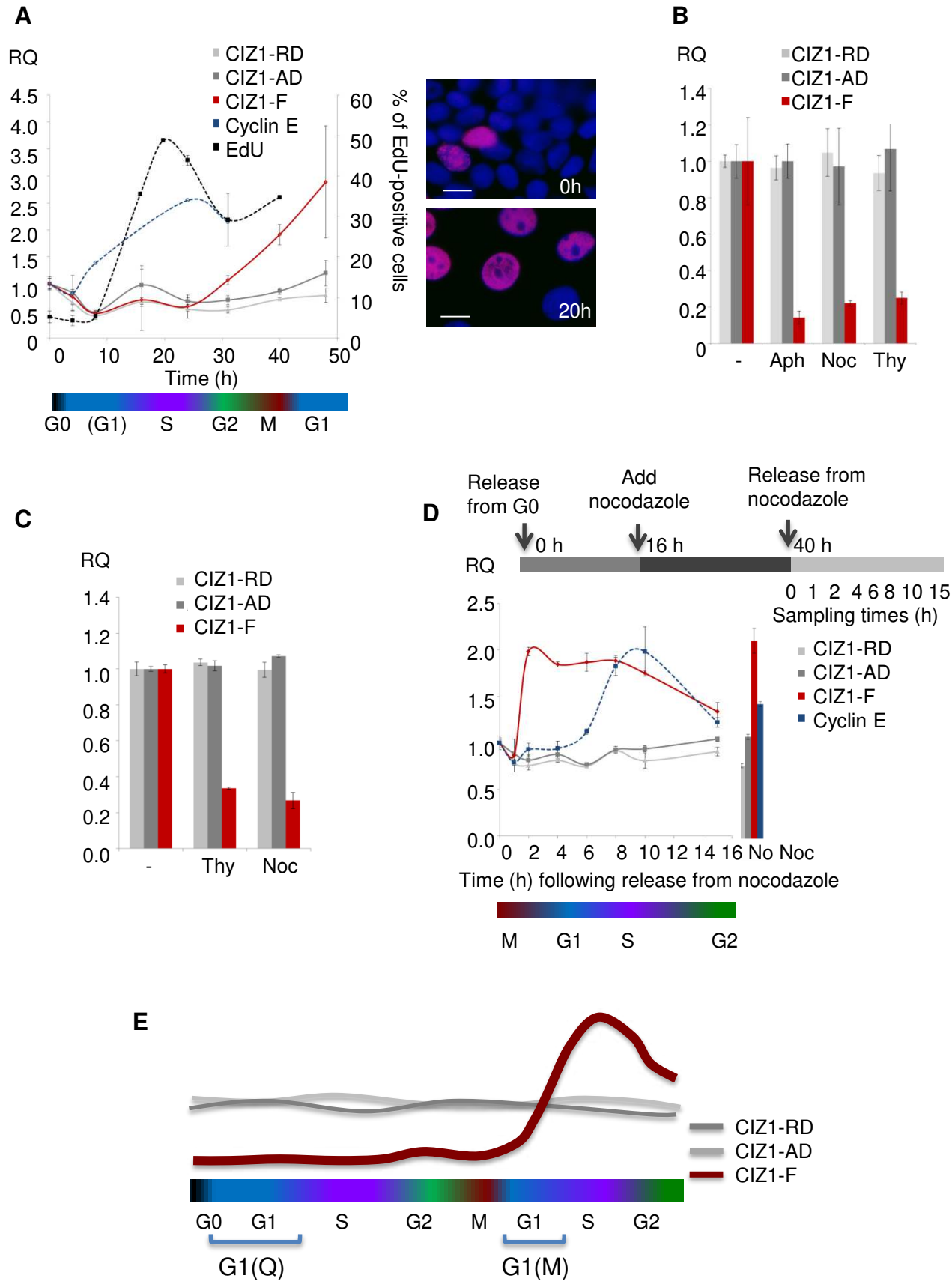

A
RQ
CIZ1-RD
CIZ1-AD
CIZ1-F
Cyclin E
EdU
Time (h)
G0 (G1) S G2 M G1
% of EdU-positive cells
0h
20h
B
RQ
CIZ1-RD
CIZ1-AD
CIZ1-F
- Aph Noc Thy
C
RQ
CIZ1-RD
CIZ1-AD
CIZ1-F
- Thy Noc
D
Release from G0
Add nocodazole
Release from nocodazole
0 h
16 h
40 h
RQ
0 1 2 4 6 8 10 15
Sampling times (h)
CIZ1-RD
CIZ1-AD
CIZ1-F
Cyclin E
0 2 4 6 8 10 12 14 16 No Noc
Time (h) following release from nocodazole
M G1 S G2
E
CIZ1-RD
CIZ1-AD
CIZ1-F
G0 G1 S G2 M G1 S G2
G1(Q)
G1(M)

Figure 5

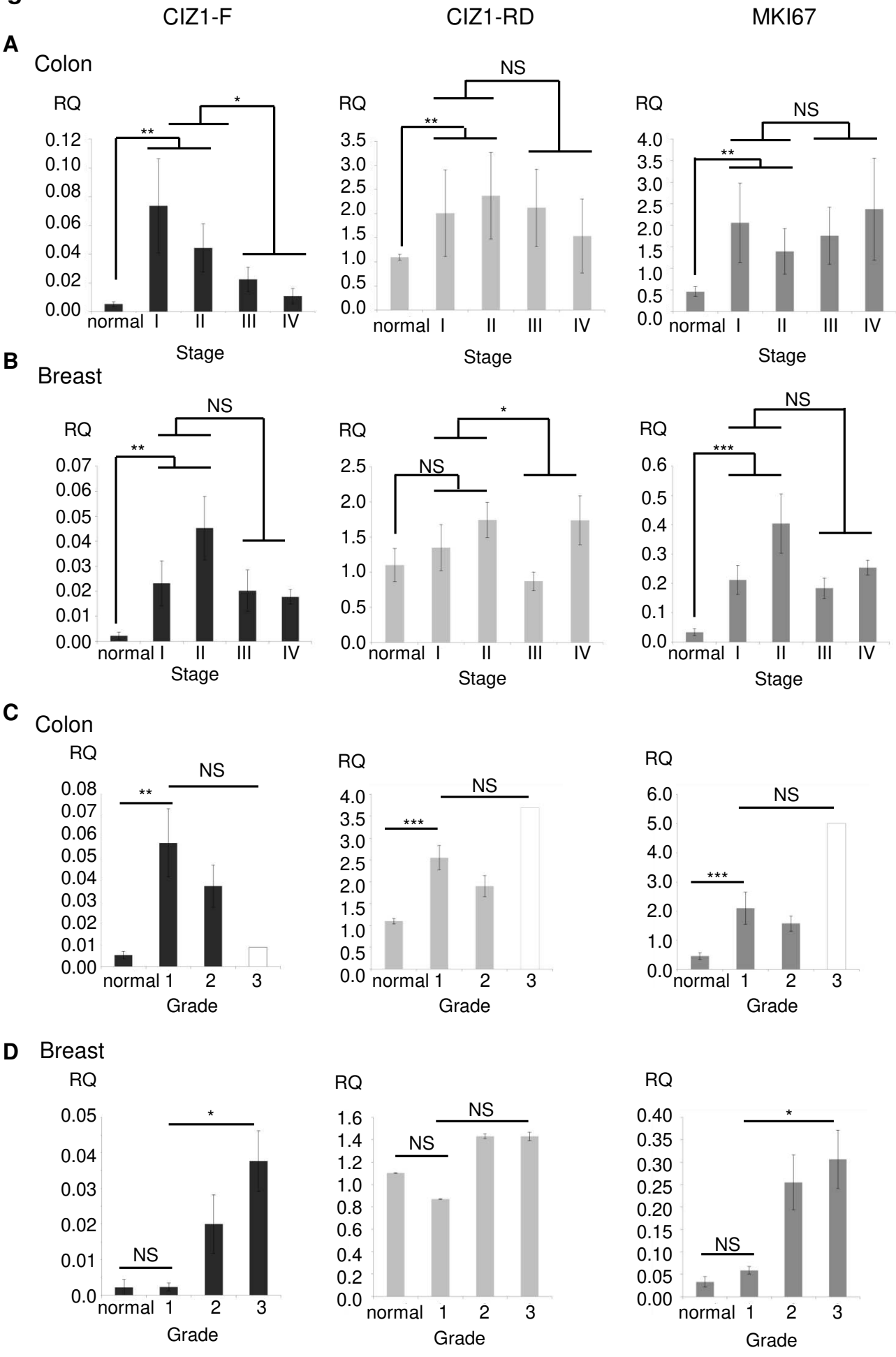

CIZ1-F
CIZ1-RD
MKI67
A
Colon
RQ
NS
**
*
normal I II III IV
Stage
B
Breast
C
Colon
Grade
D
Breast

Figure 6

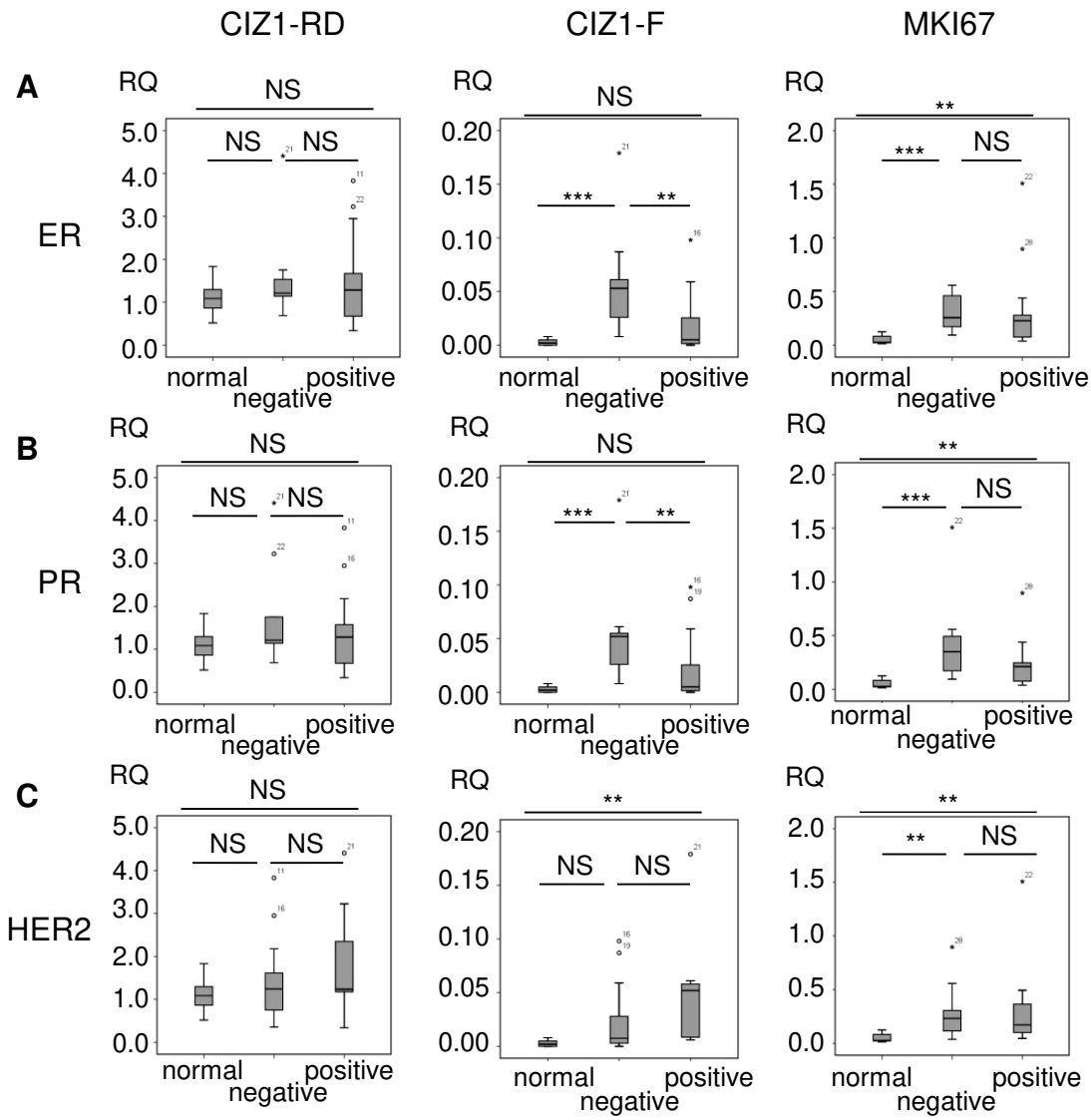
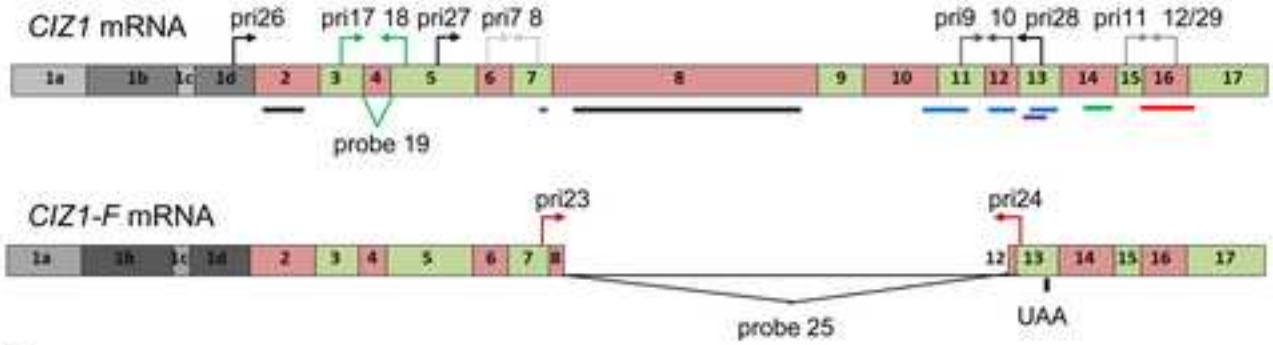

CIZ1-RD
CIZ1-F
MKI67
A
ER
B
PR
C
HER2
RQ
NS
**
***
normal
positive
negative

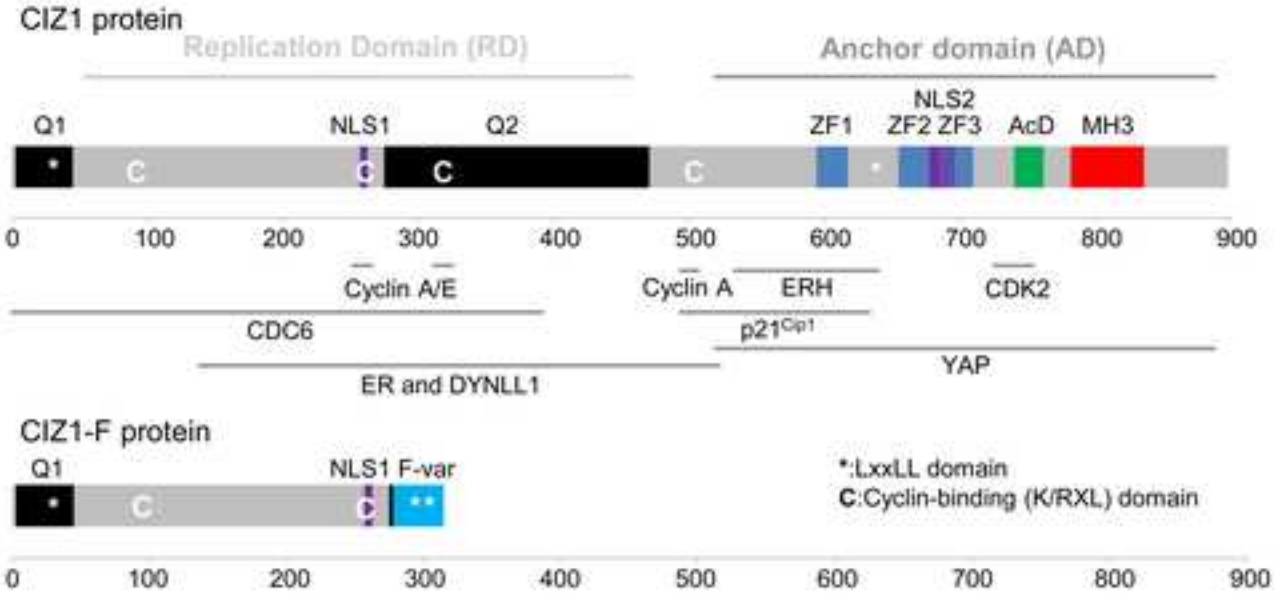
Figure 1_hres

Figure 1

A



B



C

MFSQQQQQQLQQQQQQLQQLQQQQLQQQQLQQQQLLQLQQLLQQSPPQ
APLPMAVSRGLPPQQPQQPLLNLQGTNSASLLNGSMLQRALLLQQLQGLD
QFAMPPATYDTAGLTMPTATLGNLRGYGMASPGLAAPSLTPPQLATPNLQ
QFFPQATRQSLLGPPPVGVPMNPSQFNLSGRNPQKQARTSSSTTPNRKDS
SSQTMPVEDKSDPPEGSEEAAEPRMDTPEDQDLPPCPEDIAKEKRTPAPE
PEPCEASELPAKRLRSSEEPTEKEPPGQLQDTGPQDCQTILATLLHRLQPL
LQNPSQVCGAREVPGA

Figure 2_hres

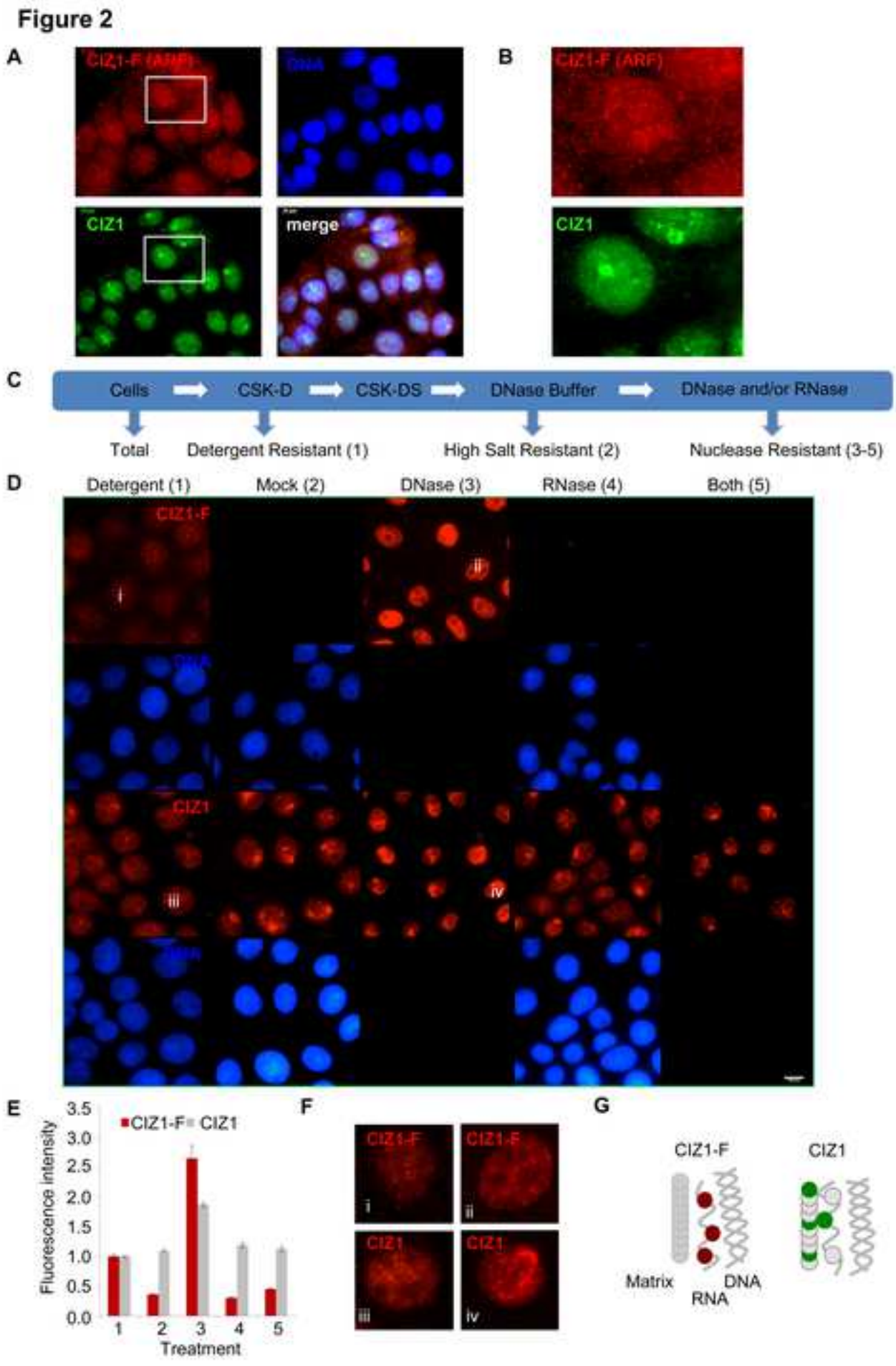


Figure 3_hres

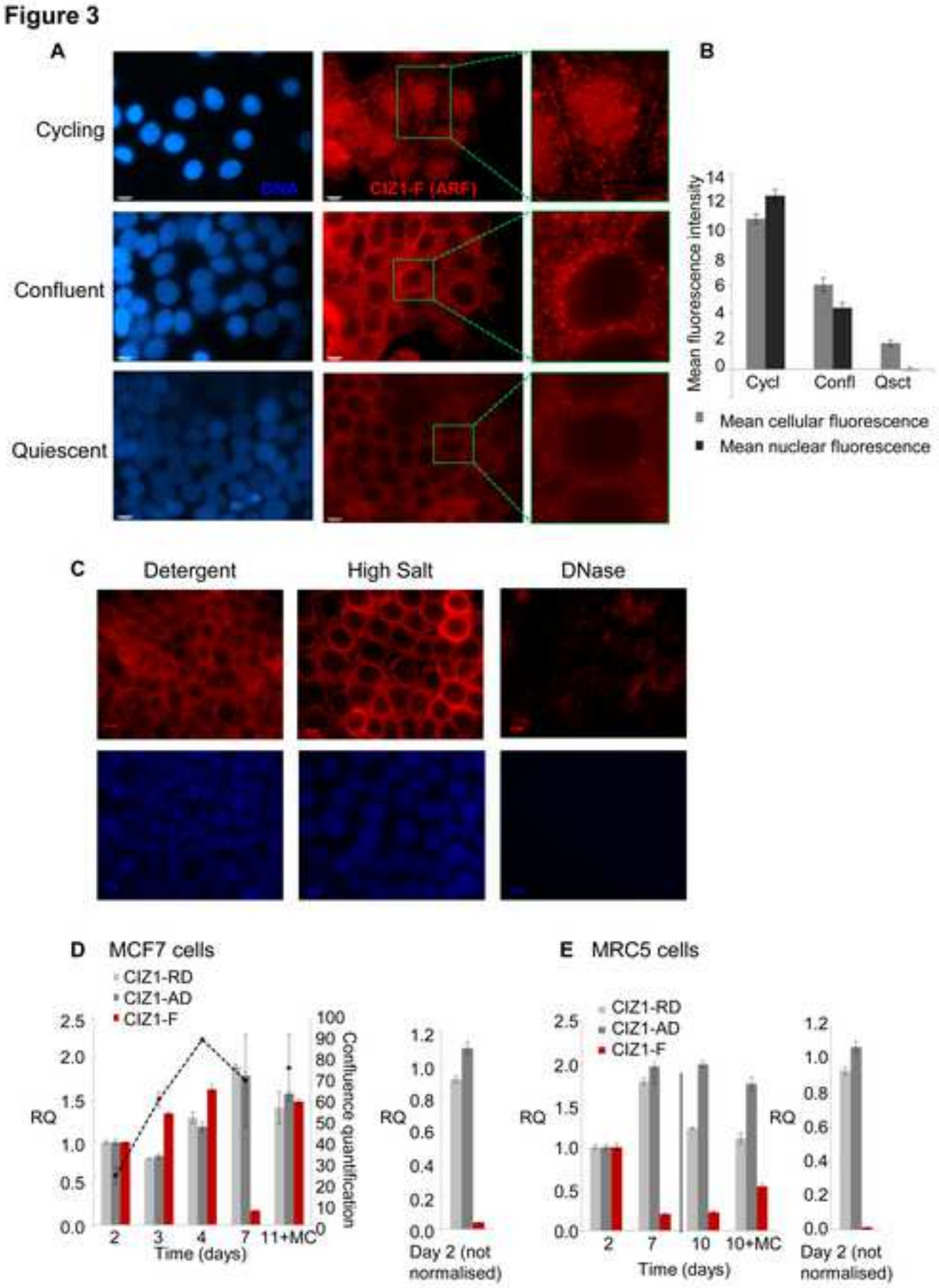


Figure 4_hres

Figure 4

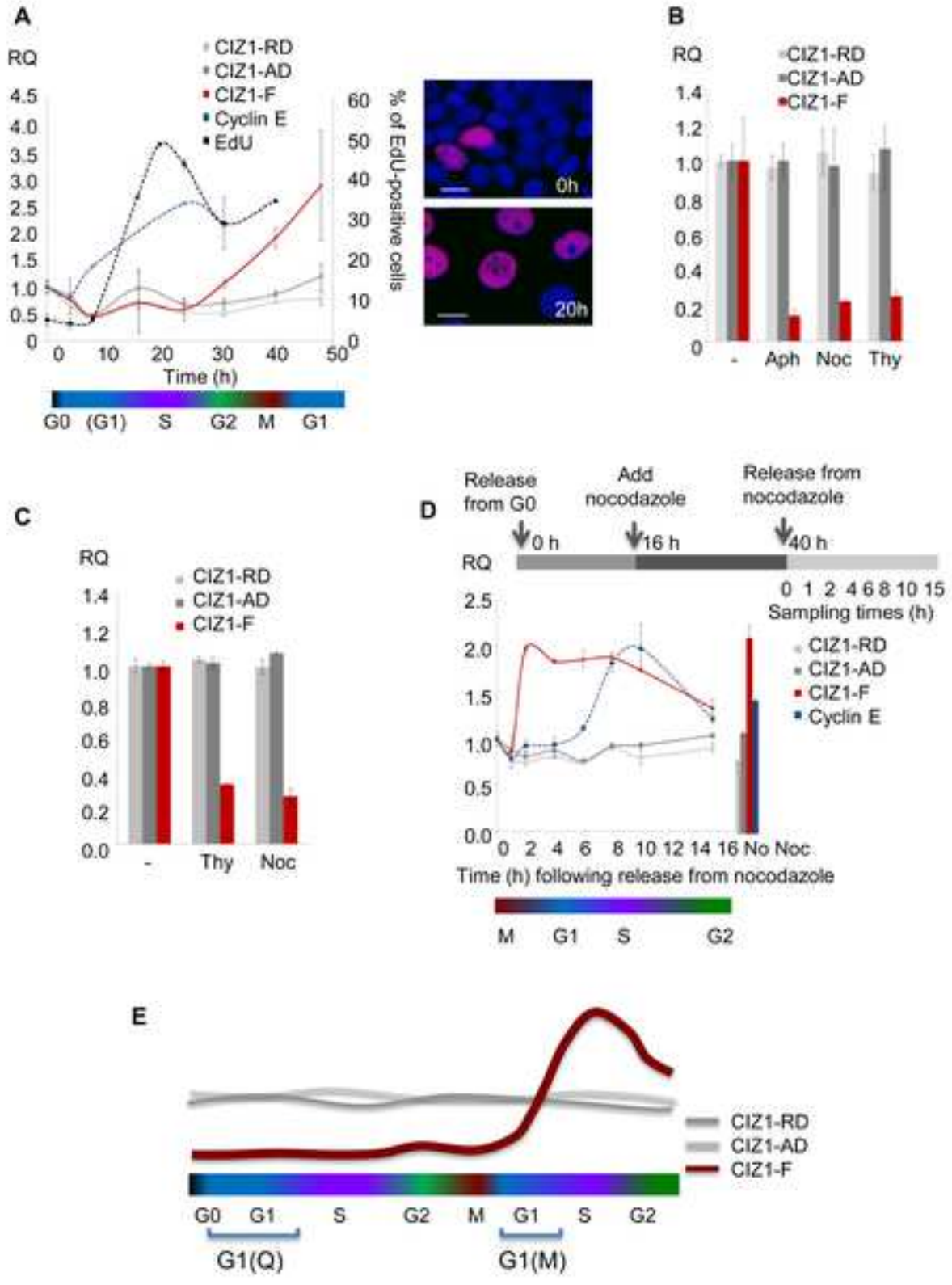


Figure 5_hres

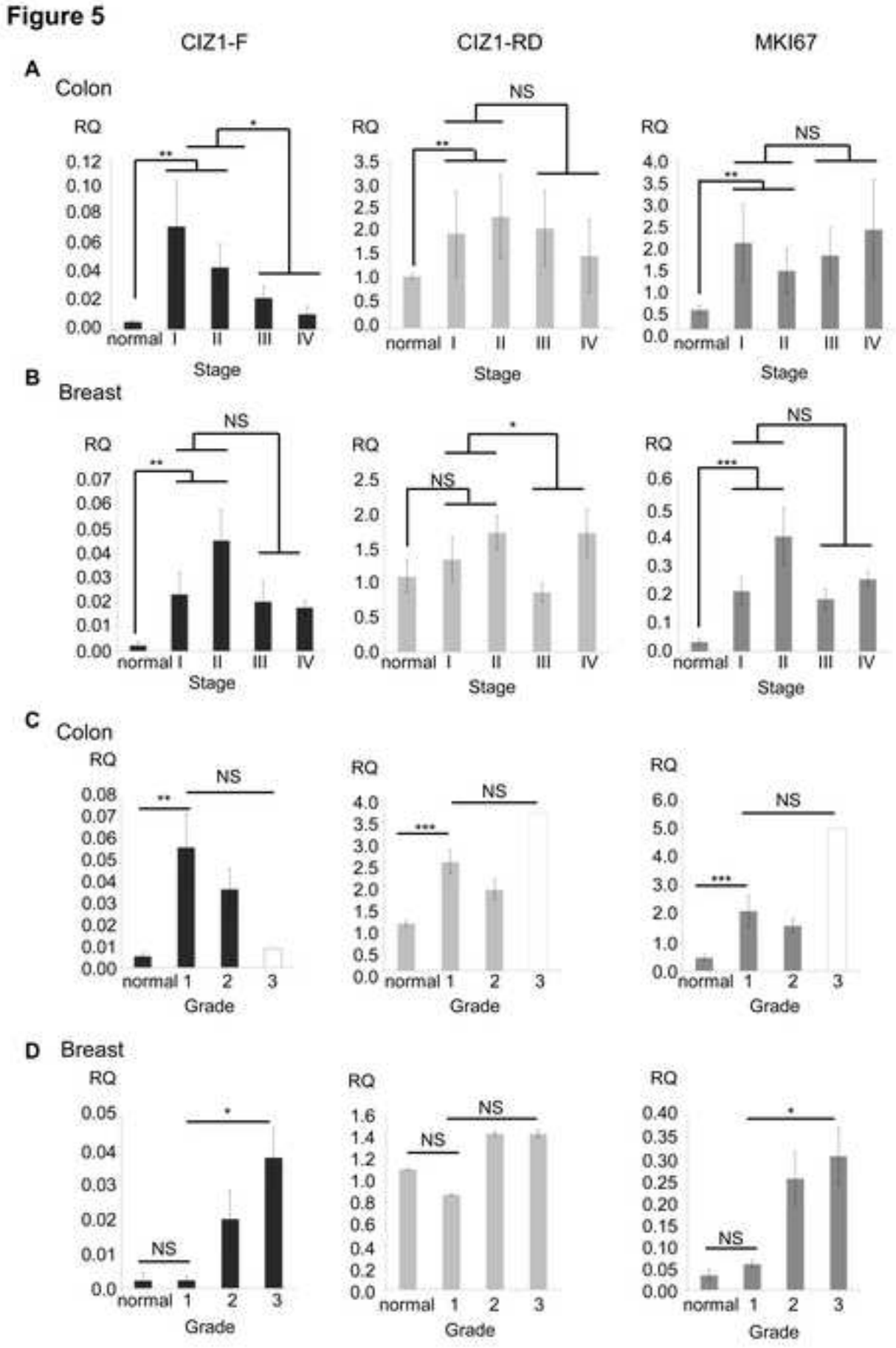
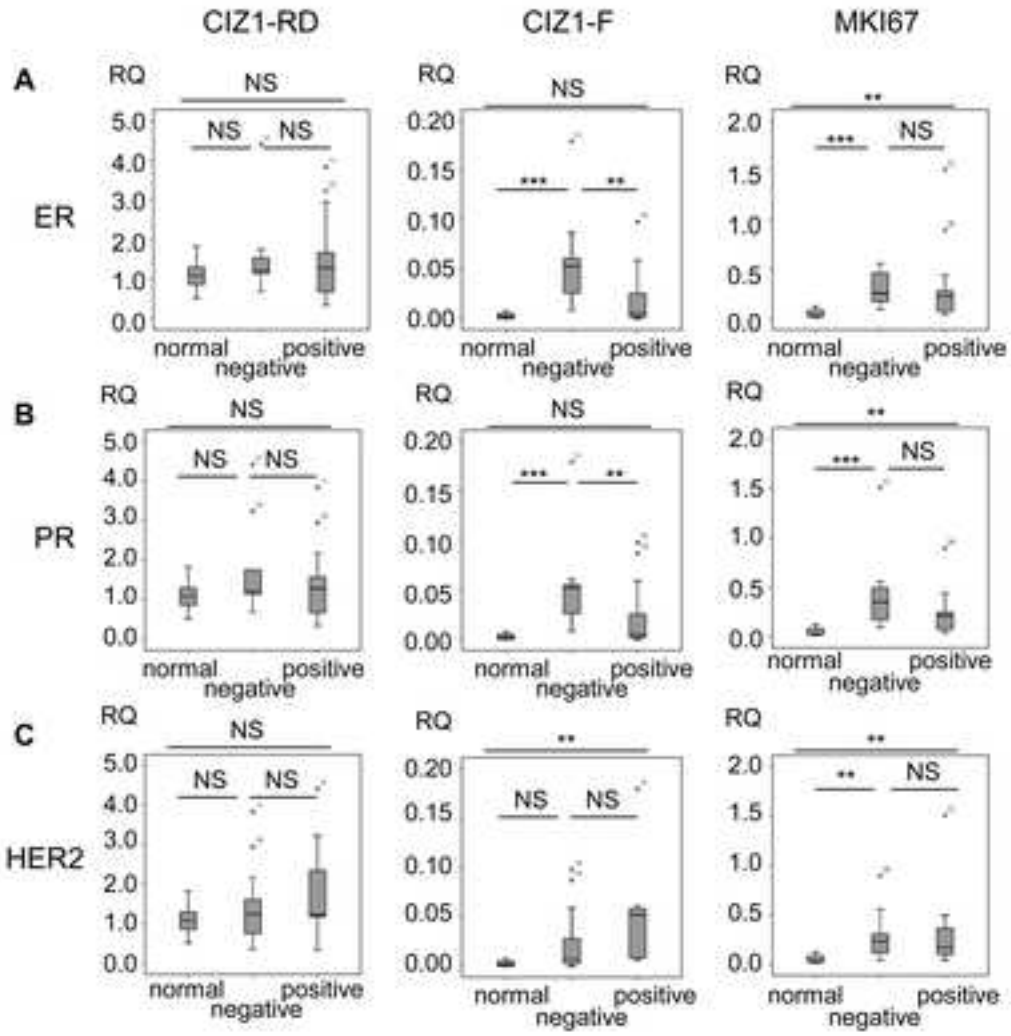
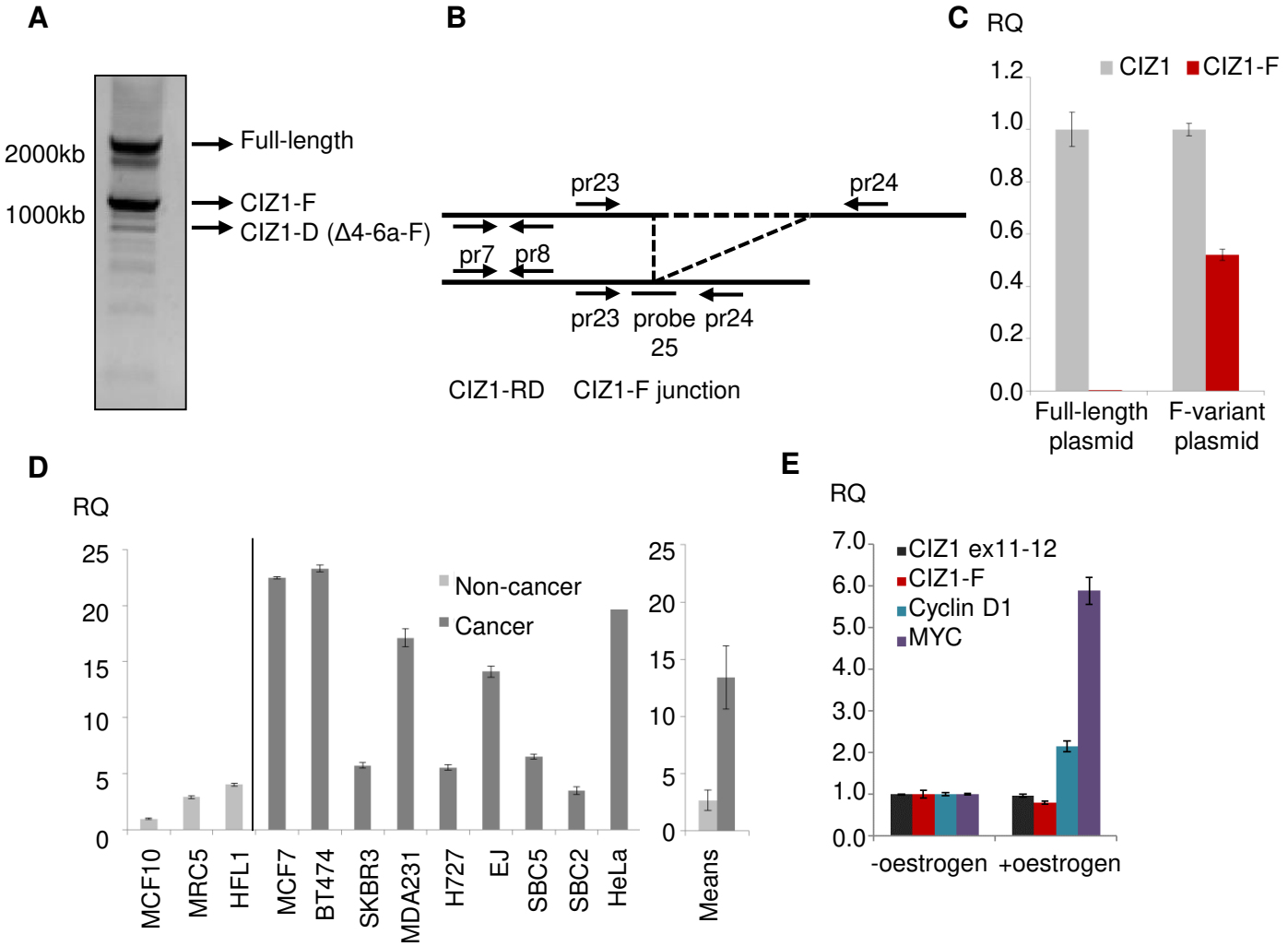


Figure 6_hres

Figure 6



Supplementary Figure 1



Supplementary Figure 1. Validation of quantitative RT-PCR tools, *CIZ1*-F expression in cell lines and in relation to oestrogen.

(A) PCR products generated using cDNA from cycling MCF-7 cells, using a forward primer in exon 1d and a reverse primer in exon 16. Major products are full-length *CIZ1* and *CIZ1*-F, but a number of smaller products are evident including *CIZ1*-F also lacking exon 4, 5 and part of 6 (*CIZ1*-D; **Supplementary Table 1**). These products were cloned and sequence verified.

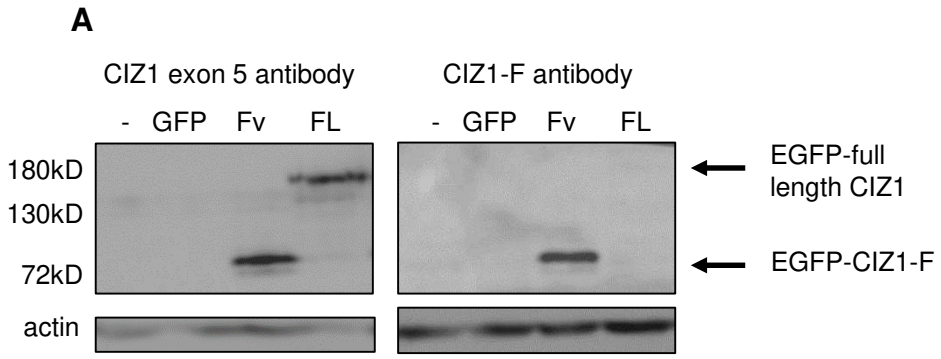
(B) Schematic showing detection tools for quantitative RT-PCR (qRT-PCR) for CIZ1 replication domain (CIZ1-RD) and CIZ1-F.

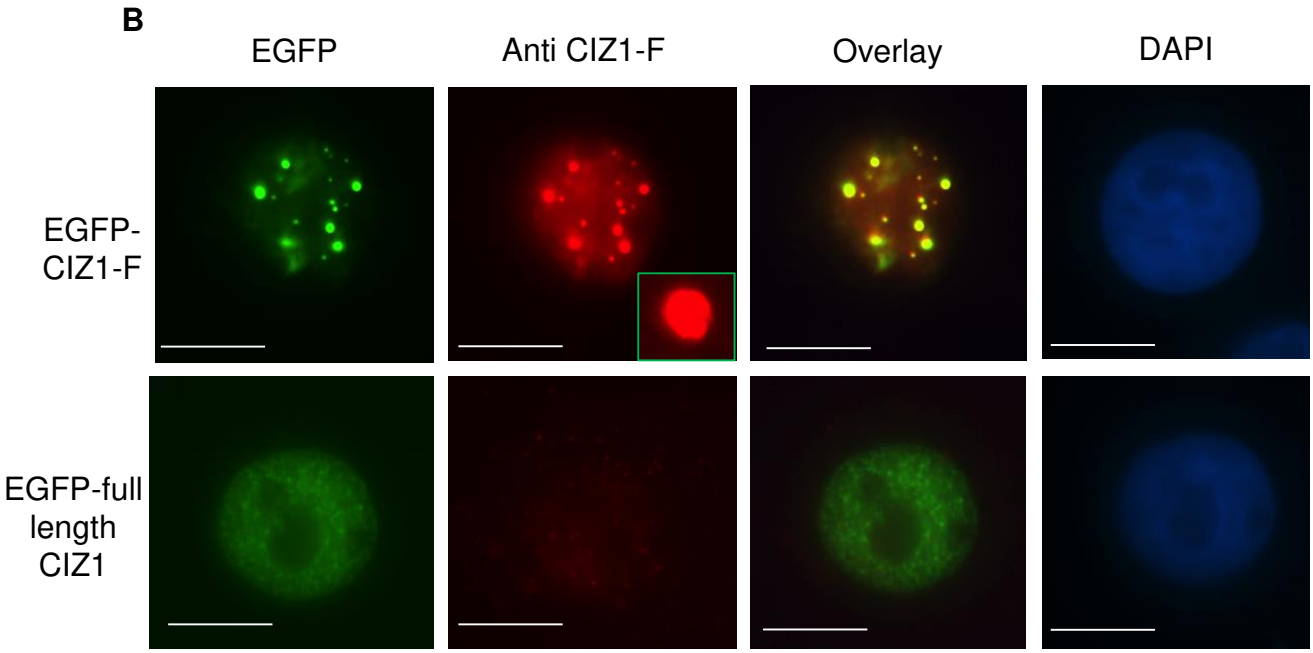
(C) Quantitative RT-PCR using full-length *CIZ1* and *CIZ1*-F plasmids as template, showing selective amplification of *CIZ1*-F template by *CIZ1*-F detection tools shown in (B). Full-length *CIZ1* plasmid is not detected by the *CIZ1*-F primer-probe set (CT-value > 40), whereas both plasmids are amplified by *CIZ1*-RD primer sets. Graphs show expression levels relative to *CIZ1*-RD ± SEM for three technical replicates.

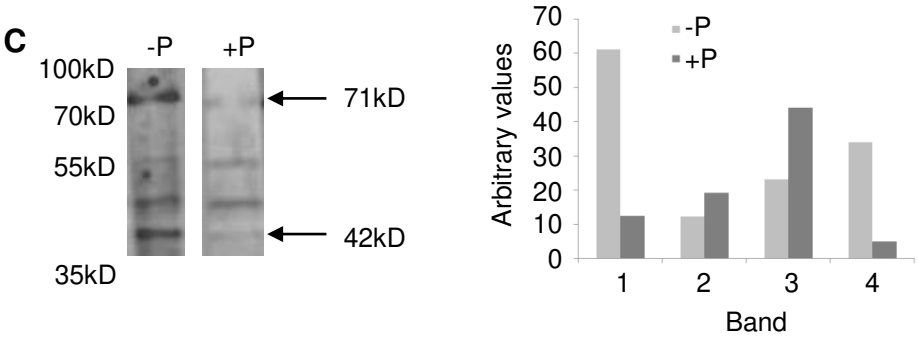
(D) Left: *CIZ1*-F transcript in a set of cancer-derived cell lines (dark grey) and three non-cancer comparators (light grey). Relative quantities (RQ) are shown ± SEM for three technical replicates, expressed relative to MCF-10A normal breast epithelial cells which is set at 1. All data is normalised to *ACTB* and *CYPA*. Right: mean value of the normal cell lines compared to the mean value of the cancer cell lines ± SEM. Material from cell lines (EJ, H727, HeLa, HFL1, SBC2, SBC5) other than those indicated in the Methods section was available from previous studies.[1, 2] Additional RNA from breast cancer cell lines BT474, SKBR3 and MDA-MB-231 was kindly provided by Dr. William Brackenbury, Department of Biology, The University of York, York, United Kingdom.

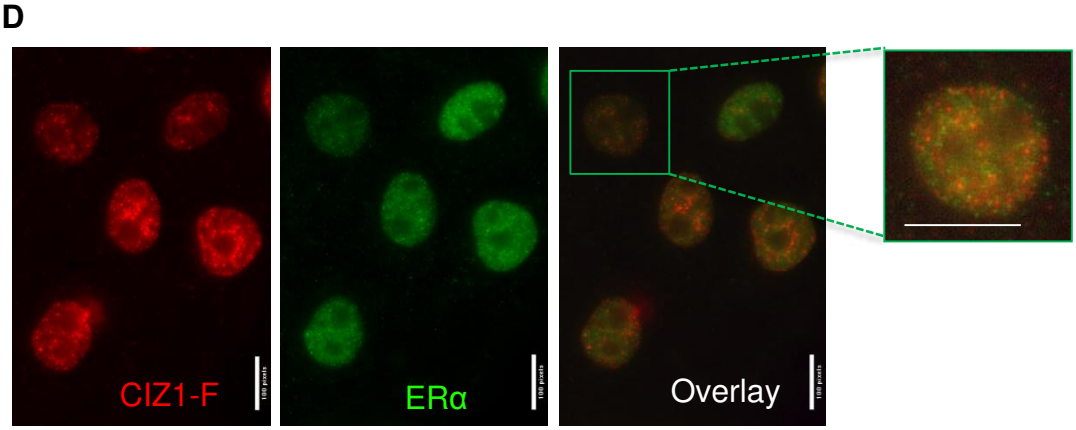
(E) Effect of oestrogen (1 nM for 16 h) on *CIZ1*-F and full-length *CIZ1* mRNA expression levels in MCF-7 cells. Oestrogen-responsive genes *CCND1* and *MYC* are shown for comparison. Mean levels (± SEM) from three technical replicates are shown relative to the levels in untreated cells. All primer and probe sequences are in **Supplementary Table 2**.

Supplementary Figure 2



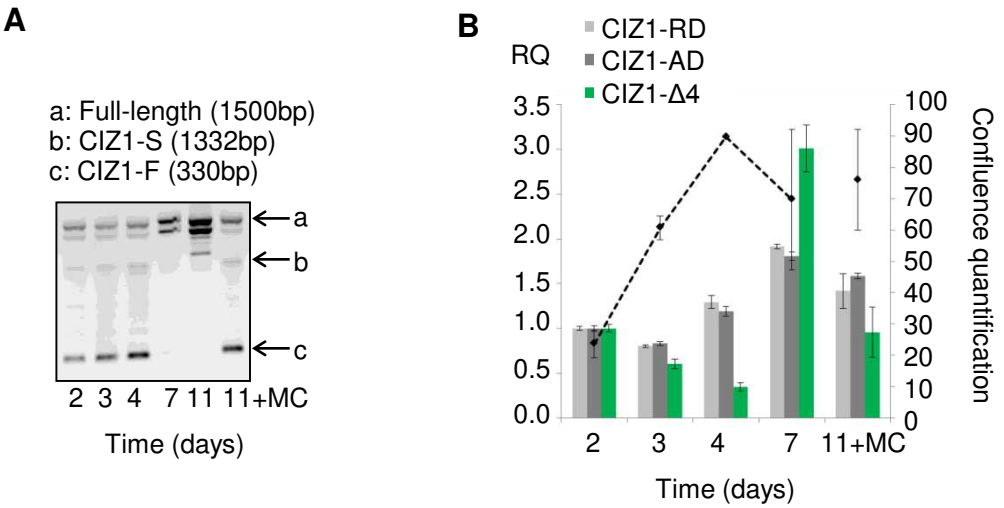






Supplementary Figure 2. Validation of CIZ1-F antibodies, and relation of CIZ1-F to ERα.
(A) Western blots showing whole cell lysates from MCF-7 cells untransfected (-), or transiently transfected with empty vector (GFP), EGFP-CIZ1-F or EGFP-full-length CIZ1. Lysates were probed with an antibody that recognizes an epitope in exon 5 (left), or affinity-purified CIZ1-F-specific polyclonal antibody (right). Actin is shown as a loading control.
(B) MCF-7 cells transfected with EGFP-full-length CIZ1 or EGFP-CIZ1-F were harvested after 24 h and washed with detergent-cytoskeletal buffer before fixation. Affinity purified CIZ1-F antibody reacted with ectopically expressed CIZ1-F but not full-length CIZ1. The image of CIZ1-F antibody on CIZ1-F-transfected cells was taken at a lower exposure time than the rest of the panel, with comparable exposure inset. Note that ectopic CIZ1-F aggregates into large foci, which is not seen with endogenous protein. Bars are 10 microns.
(C) Western blots showing whole cell lysates from MCF7 cells probed with the affinity-purified CIZ1-F -specific polyclonal antibody with (+P) or without (-P) blocking peptides. The panel on the right shows quantifications of the four bands detected on the Western blots (1~71kD, 2~55kD, 3~46kD, 4~42kD) after background subtraction.
(D) Immunolocalization of CIZ1-F and ERα in MCF-7 cells after NM extraction using detergent-containing cytoskeleton buffer, followed by 0.5 M NaCl, then DNase. The overlay image does not support CIZ1-F /ERα colocalisation (see enhanced nucleus, boxed). Bar is 10 microns.

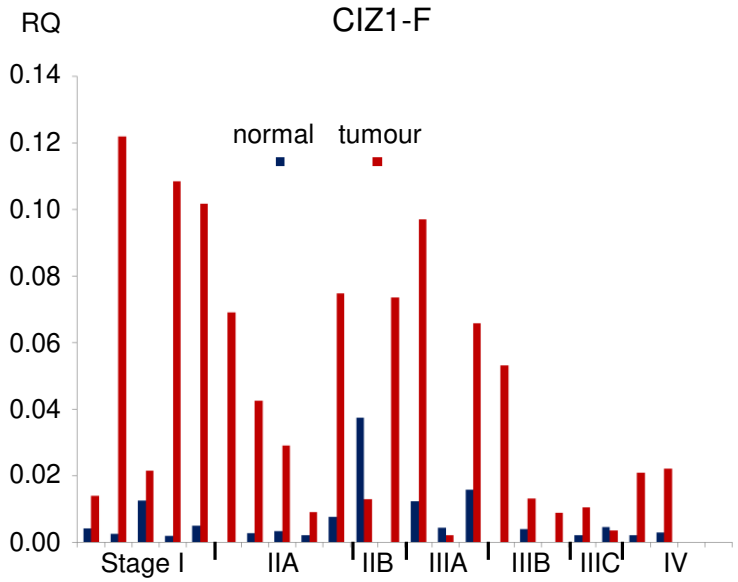
Supplementary Figure 3



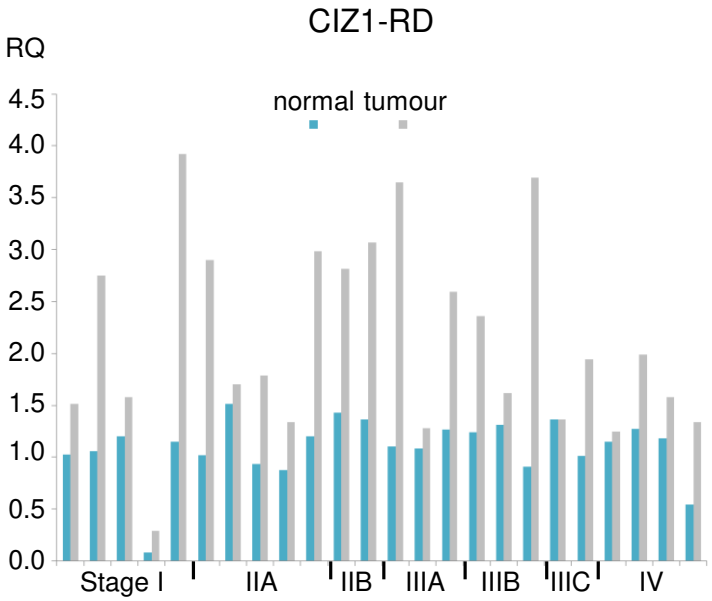
Supplementary Figure 3. *CIZ1* variant expression under different cell growth conditions.
(A) PCR products generated with a forward primer in exon 5 and a reverse primer in exon 13, using cDNA from MCF-7 cells at the same 2, 3, 4, 7 day and media changes (MC) time points as in **Fig. 3D**. Full-length *CIZ1*, *CIZ1*-S and *CIZ1*-F products were cloned and sequence-verified. *CIZ1*-F peaks at 4 days and is absent by day 7, while full-length *CIZ1* continues to increase.
(B) *CIZ1*-Δ4 transcript levels in MCF-7 breast epithelial carcinoma cells at the indicated number of days post-plating, measured by qRT-PCR, in comparison with *CIZ1*-RD and *CIZ1*-AD transcript levels (see also **Fig. 3D**). A parallel culture that received regular changes of media (indicated as 'MC'), harvested at 11 days, is shown for comparison. Histograms show the mean of three technical replicates ± SEM (black dotted line). Data is expressed as relative quantification (RQ) after normalization to *ACTB* and *CYPA*, and is calibrated to levels at day 2. Cell counts at the time of harvesting are provided ± SEM (n=3). Primer and probe sequences are in **Supplementary Table 2**.

Supplementary Figure 4

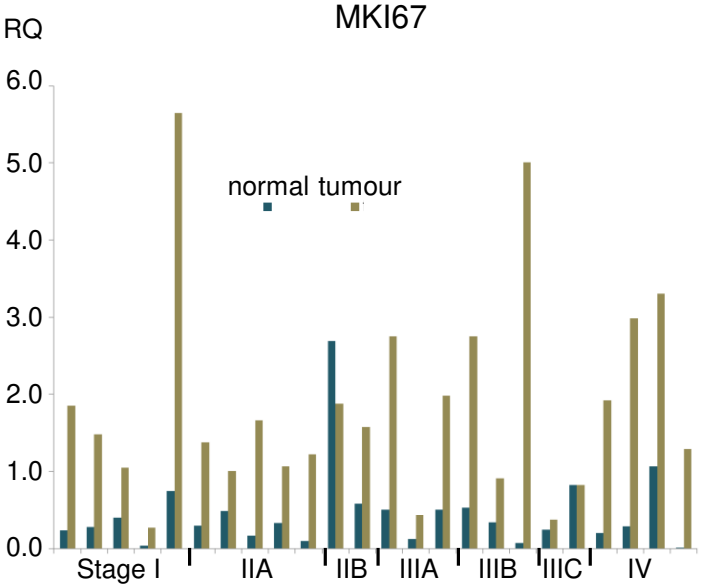
A



B

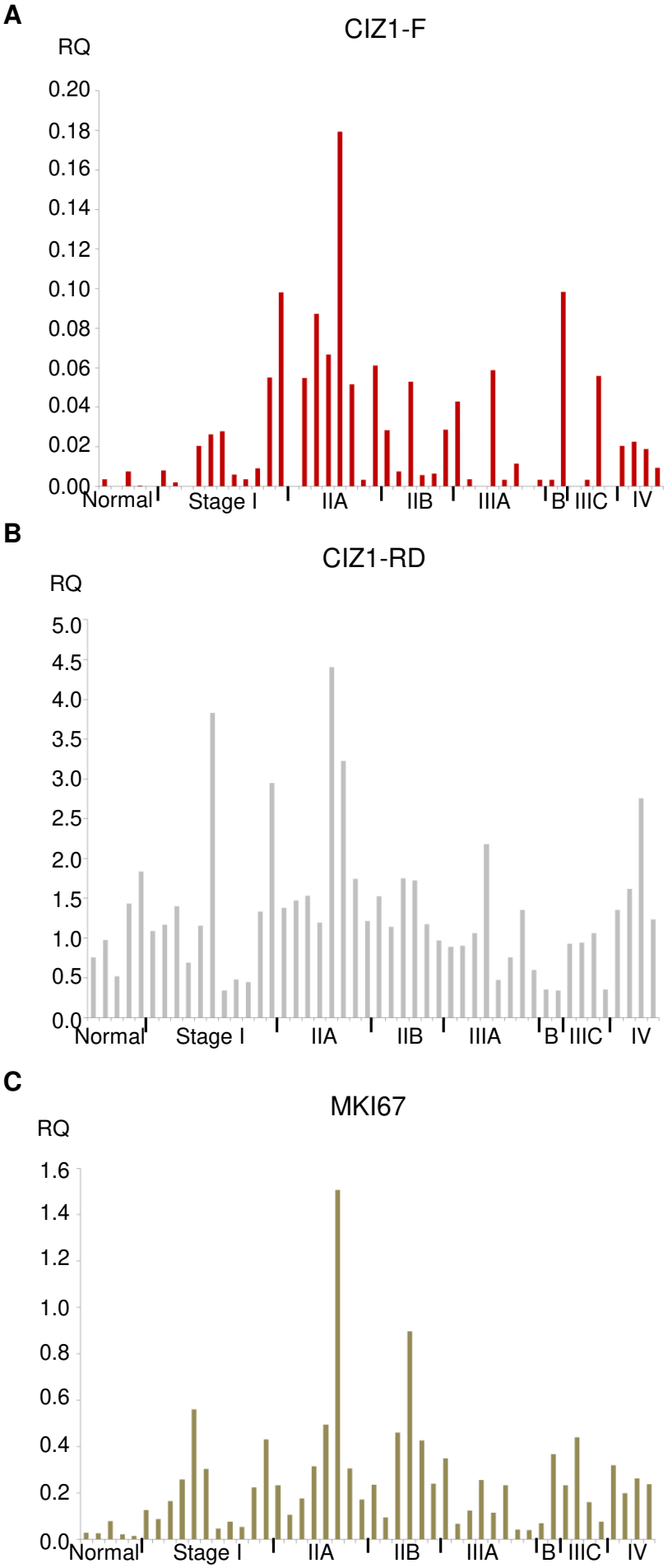


C



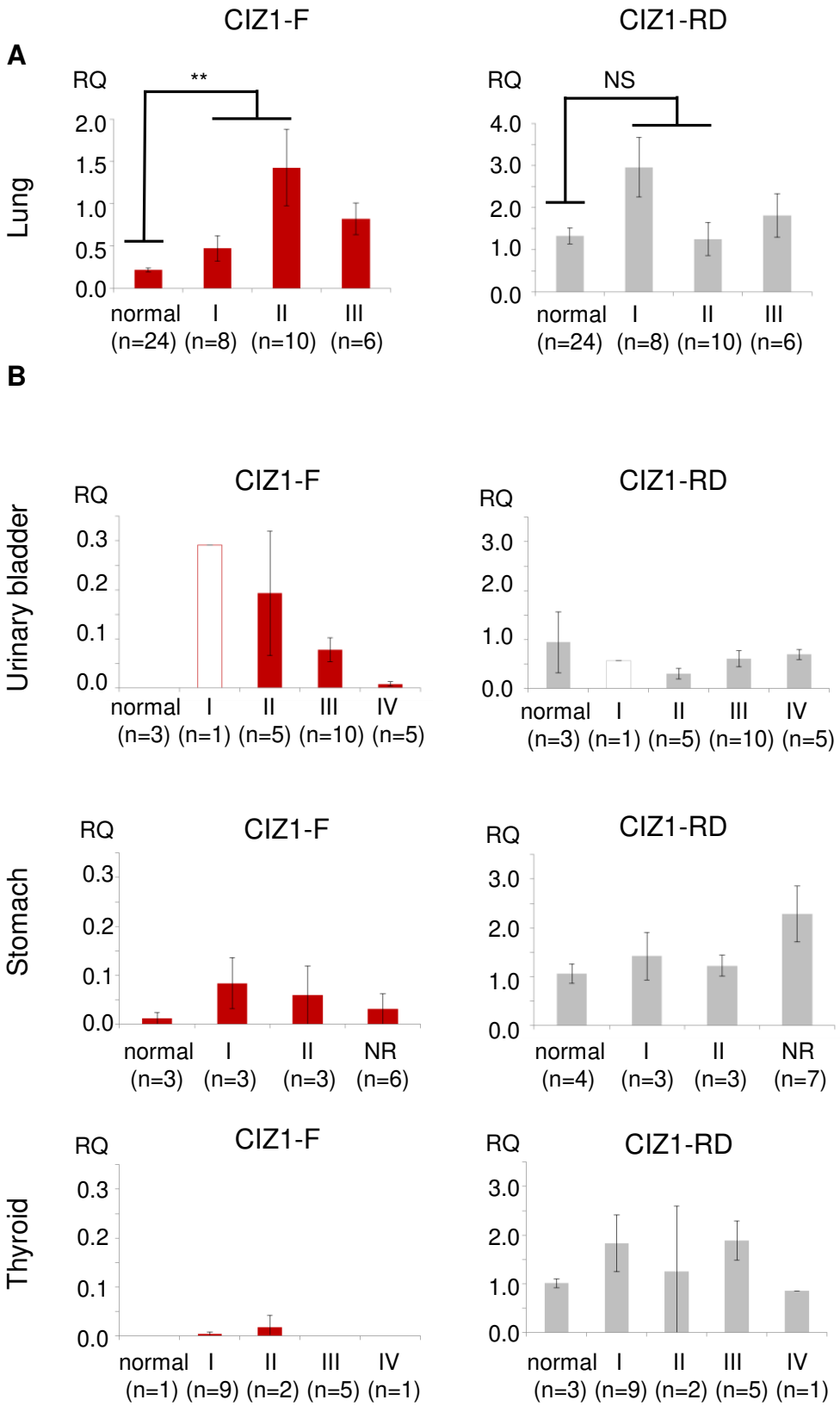
Supplementary Figure 4. *CIZ1*-F expression in human colon cancer.
Expression levels of *CIZ1*-F (A), *CIZ1* replication domain (CIZ1-RD; B), and *MKI67* (C) in 24 colon cancers from different stages, plus matched normal colon tissues (Origene colon cancer cDNA array HCRT103; sample-specific information and pathology reports can be found at www.origene.com). RQ values are expressed relative to the mean expression of *CIZ1*-RD in normal samples. Primer and probe sequences are in **Supplementary Table 2**.

Supplementary Figure 5



Supplementary Figure 5. CIZ1-F expression in human breast cancer.
Expression levels of CIZ1-F (A), *CIZ1* replication domain (*CIZ1*-RD) (B), and *MKI67* (C) in breast tissue-derived cDNAs of 5 normal samples and 43 tumours of the indicated stages (Origene breast cancer cDNA array BCRT102; sample-specific information and pathology reports can be found at www.origene.com), normalised to the mean expression of *CIZ1*-RD in the 5 normal samples. Primer and probe sequences are in **Supplementary Table 2**.

Supplementary Figure 6



Supplementary Figure 6. *CIZ1*-F expression in other cancers.
(A) Mean expression levels of *CIZ1*-F (red), *CIZ1*-RD (grey), in normal lung tissue and lung tumour samples grouped by stage. The *CIZ1*-F data for individual samples (Origene lung cancer Tissue Scan cDNA array HLRT504). RQ values are relative to the mean of RD in the normal samples. Significant differences are indicated (NS, not significant; Related-Samples Wilcoxon Signed Rank Test). There was a significant difference between normal samples and early stage cancers (I and II) for *CIZ1*-F but not *CIZ1*-RD.
(B) *CIZ1*-F (red) and *CIZ1*-RD (grey) expression levels in urinary bladder, stomach and thyroid cancer samples, compared to normal bladder, stomach and thyroid tissue in Origene cDNA array CSRT103 (sample-specific information and pathology reports can be found at www.origene.com). RQ is expressed relative to the mean *CIZ1*-RD levels of the normal samples for each cancer type. White bars indicate that only one sample was available within a category. NR, not reported. Primers and probes are listed in **Supplementary Table 2**, except for the lung array CIZ1-F data, which have been described previously.[3]

Supplementary Figure Legends

Supplementary Figure 1. Validation of quantitative RT-PCR tools, *CIZ1*-F expression in cell lines and in relation to oestrogen.

(A) PCR products generated using cDNA from cycling MCF-7 cells, using a forward primer in exon 1d and a reverse primer in exon 16. Major products are full-length *CIZ1* and *CIZ1*-F, but a number of smaller products are evident including *CIZ1*-F also lacking exon 4, 5 and part of 6 (*CIZ1*-D; **Supplementary Table 1**). These products were cloned and sequence verified.

(B) Schematic showing detection tools for quantitative RT-PCR (qRT-PCR) for CIZ1 replication domain (CIZ1-RD) and CIZ1-F.

(C) Quantitative RT-PCR using full-length *CIZ1* and *CIZ1*-F plasmids as template, showing selective amplification of *CIZ1*-F template by *CIZ1*-F detection tools shown in (B). Full-length *CIZ1* plasmid is not detected by the *CIZ1*-F primer-probe set (CT-value > 40), whereas both plasmids are amplified by *CIZ1*-RD primer sets. Graphs show expression levels relative to *CIZ1*-RD ± SEM for three technical replicates.

(D) Left: *CIZ1*-F transcript in a set of cancer-derived cell lines (dark grey) and three non-cancer comparators (light grey). Relative quantities (RQ) are shown ± SEM for three technical replicates, expressed relative to MCF-10A normal breast epithelial cells which is set at 1. All data is normalised to *ACTB* and *CYPA*. Right: mean value of the normal cell lines compared to the mean value of the cancer cell lines ± SEM. Material from cell lines (EJ, H727, HeLa, HFL1, SBC2, SBC5) other than those indicated in the Methods section was available from previous studies.[1, 2] Additional RNA from breast cancer cell lines BT474, SKBR3 and MDA-MB-231 was kindly provided by Dr. William Brackenbury, Department of Biology, The University of York, York, United Kingdom.

(E) Effect of oestrogen (1 nM for 16 h) on *CIZ1*-F and full-length *CIZ1* mRNA expression levels in MCF-7 cells. Oestrogen-responsive genes *CCND1* and *MYC* are shown for comparison. Mean levels (± SEM) from three technical replicates are shown relative to the levels in untreated cells. All primer and probe sequences are in **Supplementary Table 2**.

Supplementary Figure 2. Validation of CIZ1-F antibodies, and relation of CIZ1-F to ERα.

(A) Western blots showing whole cell lysates from MCF-7 cells untransfected (-), or transiently transfected with empty vector (GFP), EGFP-CIZ1-F or EGFP-full-length CIZ1. Lysates were probed with an antibody that recognizes an epitope in exon 5 (left), or affinity-purified CIZ1-F-specific polyclonal antibody (right). Actin is shown as a loading control.

(B) MCF-7 cells transfected with EGFP-full-length CIZ1 or EGFP-CIZ1-F were harvested after 24 h and washed with detergent-cytoskeletal buffer before fixation. Affinity purified CIZ1-F antibody reacted with ectopically expressed CIZ1-F but not full-length CIZ1. The image of CIZ1-F antibody on CIZ1-F-transfected cells was taken at a lower exposure time than the rest of the panel, with comparable exposure inset. Note that ectopic CIZ1-F aggregates into large foci, which is not seen with endogenous protein. Bars are 10 microns.

(C) Western blots showing whole cell lysates from MCF7 cells probed with the affinity-purified CIZ1-F -specific polyclonal antibody with (+P) or without (-P) blocking peptides. The panel on the right shows quantifications of the four bands detected on the Western blots (1~71kD, 2~55kD, 3~46kD, 4~42kD) after background subtraction.

(D) Immunolocalization of CIZ1-F and ERα in MCF-7 cells after NM extraction using detergent-containing cytoskeleton buffer, followed by 0.5 M NaCl, then DNase. The overlay image does not support CIZ1-F /ERα colocalisation (see enhanced nucleus, boxed). Bar is 10 microns.

Supplementary Figure 3. *CIZ1* variant expression under different cell growth conditions.

(A) PCR products generated with a forward primer in exon 5 and a reverse primer in exon 13, using cDNA from MCF-7 cells at the same 2, 3, 4, 7 day and media changes (MC) time points as in **Fig. 3D**. Full-length *CIZ1*, *CIZ1*-S and *CIZ1*-F products were cloned and sequence-verified. *CIZ1*-F peaks at 4 days and is absent by day 7, while full-length *CIZ1* continues to increase.

(B) *CIZ1*-Δ4 transcript levels in MCF-7 breast epithelial carcinoma cells at the indicated number of days post-plating, measured by qRT-PCR, in comparison with *CIZ1*-RD and *CIZ1*-AD transcript levels (see also **Fig. 3D**). A parallel culture that received regular changes of media (indicated as 'MC'), harvested at 11 days, is shown for comparison. Histograms show the mean of three technical replicates ± SEM (black dotted line). Data is expressed as relative quantification (RQ) after normalization to *ACTB* and *CYPA*, and is calibrated to levels at day 2. Cell counts at the time of harvesting are provided ± SEM (n=3). Primer and probe sequences are in **Supplementary Table 2**.

Supplementary Figure 4. *CIZ1*-F expression in human colon cancer.

Expression levels of *CIZ1*-F (A), *CIZ1* replication domain (CIZ1-RD; B), and *MKI67* (C) in 24 colon cancers from different stages, plus matched normal colon tissues (Origene colon cancer cDNA array HCRT103; sample-specific information and pathology reports can be found at www.origene.com). RQ values are expressed relative to the mean expression of *CIZ1*-RD in normal samples. Primer and probe sequences are in **Supplementary Table 2**.

Supplementary Figure 5. CIZ1-F expression in human breast cancer.

Expression levels of CIZ1-F (A), *CIZ1* replication domain (*CIZ1*-RD) (B), and *MKI67* (C) in breast tissue-derived cDNAs of 5 normal samples and 43 tumours of the indicated stages (Origene breast cancer cDNA array BCRT102; sample-specific information and pathology reports can be found at www.origene.com), normalised to the mean expression of *CIZ1*-RD in the 5 normal samples. Primer and probe sequences are in **Supplementary Table 2**.

Supplementary Figure 6. *CIZ1*-F expression in other cancers.

(A) Mean expression levels of *CIZ1*-F (red), *CIZ1*-RD (grey), in normal lung tissue and lung tumour samples grouped by stage. The *CIZ1*-F data for individual samples (Origene lung cancer Tissue Scan cDNA array HLRT504). RQ values are relative to the mean of RD in the normal samples. Significant differences are indicated (NS, not significant; Related-Samples Wilcoxon Signed Rank Test). There was a significant difference between normal samples and early stage cancers (I and II) for *CIZ1*-F but not *CIZ1*-RD.

(B) *CIZ1*-F (red) and *CIZ1*-RD (grey) expression levels in urinary bladder, stomach and thyroid cancer samples, compared to normal bladder, stomach and thyroid tissue in Origene cDNA array CSRT103 (sample-specific information and pathology reports can be found at www.origene.com). RQ is expressed relative to the mean *CIZ1*-RD levels of the normal samples for each cancer type. White bars indicate that only one sample was available within a category. NR, not reported. Primers and probes are listed in **Supplementary Table 2**, except for the lung array CIZ1-F data, which have been described previously.[3]

References

1. Higgins G, Roper KM, Watson IJ, Blackhall FH, Rom WN, Pass HI, Ainscough JF, Coverley D. Variant Ciz1 is a circulating biomarker for early-stage lung cancer. Proc Natl Acad Sci U S A 2012; 109:E3128-35.

2. Rahman F, Ainscough JF, Copeland N, Coverley D. Cancer-associated missplicing of exon 4 influences the subnuclear distribution of the DNA replication factor CIZ1. Hum Mutat 2007; 28:993-1004.

3. Rahman FA, Aziz N, Coverley D. Differential detection of alternatively spliced variants of Ciz1 in normal and cancer cells using a custom exon-junction microarray. BMC Cancer 2010; 10:482.

Supplementary Tables

Supplementary Table 1 - Overview of common CIZ1-variants*							
Name	**Splicing event**	**Expression in normal tissue**	**Expression in disease**	**Cause**	**Function**	**Subcellular expression pattern**	**References**
Single events							
Ciz1-Δ3	Deletion of exon 3 (fs; may be in combination with Δ4 and Δ6a)	In mouse testis during development	Unknown	May be developmentally regulated	Unknown	Unknown	Greaves[1]
CIZ1-Δ4 (NP-94 in combination with Δ6a)	Deletion of exon 4 (in frame)	Expressed during mouse development and massively upregulated in differentiating male germ cells	Common in Ewing tumours	Expansion of upstream intronic splicing regulator	Active in DNA replication	Fails to form subnuclear foci and has a dominant-negative effect on other CIZ1 variants	Warder & Keherly[2] Rahman,[3] Greaves[1]
CIZ1-Δ6a	Deletion of first five amino acids (DSSSQ) of exon 6	Detected in human and mouse	Detected but not linked to disease	Unknown	Unknown, possible phosphorylation site	Unknown	Warder and Keherly,[2] Greaves[1]
CIZ1-F (f-variant)	Deletion of exon 9-11 and part of exon 8 and 12 (fs)	In several cell lines; downregulated in quiescence	Overexpressed in (early stage) tumours and cancer cell lines. Lower expression in ER-positive breast cancer	Unknown	Associated with proliferation; involved in G_1	Mainly nuclear in cycling cells, but cytoplasmic in contact-inhibited cells. Part of the RNA-component of the NM	Rahman[4] and this study
CIZ1-M (m-variant)	Deletion of 84 bp from exon 8 (in frame)	Very low expression levels	Unknown, low	Unknown (repetitive sequences)	Unknown	Diffuse nuclear, abrogation of the association with the NM	Dahmcke[5]
CIZ1-S (s-variant)	Deletion of 168 bp from exon 8 (in frame)	Lower in normal hippocampi	Upregulated in Alzheimer's disease	Unknown (repetitive sequences)	Unknown	Unknown	Dahmcke[5]

CIZ1-B (b-variant)	Deletion of last 24 nt of exon 14 (in frame)	Low expression levels	Prevalent in lung cancer samples	Unknown	Circulating biomarker	Diffuse nuclear	Higgins[6]
Combined events (selection)[+]							
CIZ1-Δ4-S (Δ4-s-variant)	Deletion of exon 4 and 168 bp from exon 8 (in frame)	Unknown	Unknown	Unknown	Unknown	Unknown	UniProt
CIZ1-D-F (d-f-variant)	Deletion of exon 4, 5 and first five amino acids (DSSSQ) of exon 6 and the CIZ1-F event (fs)	Unknown	Expressed in MCF-7 breast cancer cells	Unknown	Fs may lead to the use of an alternative TSS in exon 6	Unknown	This study
CIZ1-D-S (d-s-variant)	Deletion of exon 4, 5 and DSSSQ and 168 bp from exon 8 (fs)	Unknown	Detected in MCF-7 breast cancer cells	Unknown	Fs may lead to the use of an alternative TSS in exon 6	Unknown	Our unpublished data
CIZ1-S-Δ9 (s-Δ9-variant)	Deletion of 168 bp from exon 8, and exon 9 (fs)	Unknown	Ewing tumour cell lines and H727 lung cancer cell line	Unknown	Unknown	Unknown	Rahman[4]
ECIZ1 (mouse)	Exclusion of exon 2 and part of exon 3, and exons 6 and 8	Embryonic mouse tissues	Unknown	Unknown	Active in DNA replication	Forms nuclear foci, but slightly less efficiently	Coverley,[7] Ainscough[8]

*Note that additional *CIZ1* splicing events have been detected using an exon-junction microarray,[4] which are not all listed here.

[+]Many additional combinations have been identified. For instance, almost all *CIZ1* splicing events occur with or without Δ6a.

Abbreviations used: fs, frame-shift; NM, nuclear matrix; nt, nucleotides; TSS, translational start site.

Supplementary Table 2 - Primers and probes

A) SYBR Green primers		Description	Sequence
	1	*ACTB* forward	5' CAA CCG CGA GAA GAT GAC C 3'
	2	*ACTB* reverse	5' TCC AGG GCG ACG TAG CAC A 3'
	3	*CCND1* forward	5' TCC TCT CCA AAA TGC CAG AG 3'
	4	*CCND1* reverse	5' GGC GGA TTG GAA ATG AAC TT 3'
	5	*CCNE1* forward	5' GAA ATG GCC AAA ATC GAC AG 3'
	6	*CCNE1* reverse	5' TCT TTG TCA GGT GTG GGG A 3'
	7	*CIZ1* exon 6 forward	5' TCT TCT CAG ACA ATG CCT GTG G 3'
	8	*CIZ1* exon 7 reverse	5' GCA GGG CGG TAA ATC TTG G 3'
	9	*CIZ1* exon 11 forward	5' GGA GAT CCA GCA CAT GAG CC 3'
	10	*CIZ1* exon 12 reverse	5' CAG GTC CCC CAT GTA GTA GAG CT 3'
	11	*CIZ1* exon 15 forward	5' ACC TAC AGC CCC AAT ACT GCA TAT 3'
	12	*CIZ1* exon 16 reverse	5' AGA GCT GTG CCC CTG AGT TG 3'
	13	*CYPA* forward	5' TTT CAT CTG CAC TGC CAA GAC T 3'
	14	*CYPA* reverse	5' TTC ATG CCT TCT TTC ACT TTG C 3'
	15	*MYC* forward	5' GCC ACG TCT CCA CAC ATC AG 3'
	16	*MYC* reverse	5' TCT TGG CAG CAG GAT AGT CCT T 3'
B) Taqman primers & probes		**Description**	**Sequence**
	17	*CIZ1* exon 3 forward	5' CTC CAT GCT GCA GAG AGC TT 3'
	18	*CIZ1* exon 5 reverse	5' GGC CTG GGG AAA GAA CTG TT 3'
	19	*CIZ1* Δ4 probe	5' FAM-CAG CAG TTG CAA GGT AAC CTC CGA-TAMRA 3'
	20	*CIZ1* exon 6 forward	5' TGC CTG TGG AAG ACA AGT CA 3'
	21	*CIZ1* exon 7 reverse	5' TGC TGG AGT GCG TTT TTC CT 3'
	22	*CIZ1* exon 7 probe	5' JOE CCC TGC CCA GAG GAC ATC GCC-BH 3'
	23	*CIZ1* exon 7 forward	5' CTG CCA GCA AAG AGA TTG AG 3'
	24	*CIZ1* exon 13 reverse	5' TGC GAG GGG TTT TGA AGT AG 3'
	25	*CIZ1* f-variant probe	5' FAM-CAG GGC AGT TAC AGG ACA CAG GAC-TAMRA 3'
D) PCR primers		**Description**	**Sequence**
	26	*CIZ1* exon 1d forward	5' GCG ACT TGA GCG TTG AG 3'
	27	*CIZ1* exon 5 forward	5' CTC CTC CTC TAC CAC CCC C 3'
	28	*CIZ1* exon 13 reverse	5' CAG AAG GGT CGC AAG GAT TGT 3'
	29	*CIZ1* exon 16 reverse	5' AGA GCT GTG CCC CTG AGT TG 3'

References

1. Greaves EA, Copeland NA, Coverley D, Ainscough JF. Cancer-associated variant expression and interaction of CIZ1 with cyclin A1 in differentiating male germ cells. Journal of cell science 2012; 125:2466-77.

2. Warder DE, Keherly MJ. Ciz1, Cip1 interacting zinc finger protein 1 binds the consensus DNA sequence ARYSR(0-2)YYAC. J Biomed Sci 2003; 10:406-17.

3. Rahman F, Ainscough JF, Copeland N, Coverley D. Cancer-associated missplicing of exon 4 influences the subnuclear distribution of the DNA replication factor CIZ1. Hum Mutat 2007; 28:993-1004.

4. Rahman FA, Aziz N, Coverley D. Differential detection of alternatively spliced variants of Ciz1 in normal and cancer cells using a custom exon-junction microarray. BMC Cancer 2010; 10:482.

5. Dahmcke CM, Buchmann-Moller S, Jensen NA, Mitchelmore C. Altered splicing in exon 8 of the DNA replication factor CIZ1 affects subnuclear distribution and is associated with Alzheimer's disease. Mol Cell Neurosci 2008; 38:589-94.

6. Higgins G, Roper KM, Watson IJ, Blackhall FH, Rom WN, Pass HI, Ainscough JF, Coverley D. Variant Ciz1 is a circulating biomarker for early-stage lung cancer. Proc Natl Acad Sci U S A 2012; 109:E3128-35.

7. Coverley D, Marr J, Ainscough J. Ciz1 promotes mammalian DNA replication. Journal of cell science 2005; 118:101-12.

8. Ainscough JF, Rahman FA, Sercombe H, Sedo A, Gerlach B, Coverley D. C-terminal domains deliver the DNA replication factor Ciz1 to the nuclear matrix. Journal of cell science 2007; 120:115-24.