scientific reports

Check for updates

OPEN

Cell type specific long non-coding RNA targets identified by integrative analysis of single-cell and bulk colorectal cancer transcriptomes

Ante Mihaljevic[1,2], Philip D. Rubin[1,2], Panagiotis Chouvardas[3,4,6✉] & Roberta Esposito[1,2,5,6✉]

Long non-coding RNAs (lncRNAs) represent an emerging class of genes which play significant and diverse roles in human cancers. Nevertheless, the functional repertoires of lncRNAs in cancer cell subtypes remains unknown since most studies are focused on protein coding genes. Here, we explored the contribution of lncRNAs in Colorectal Cancer (CRC) heterogeneity. We analyzed 49'436 single-cells from 29 CRC patients and showed that lncRNAs are significantly more cell type specific compared to protein-coding genes. We identified 996 lncRNAs strongly enriched in epithelial cells. Among these, 98 were found to be differentially expressed in tumor samples compared to normal controls, when integrating 270 bulk CRC profiles. We validated the upregulation of two of them (*CASC19* and *LINC00460*) in CRC cell lines and showed their involvement in CRC proliferation by CRISPR-Cas9 knock down experiments. This study highlights a list of novel RNA targets for potential CRC therapeutics, substantiated through experimental validation.

Colorectal cancer (CRC) is the third most common cancer and the second leading cause of cancer-related deaths worldwide[1]. Tumor heterogeneity, a hallmark of CRC, has emerged as a critical factor complicating effective treatment strategies. Within a single tumour mass, CRC exhibits remarkable cellular and molecular diversity, with distinct subpopulations of cancer cells exhibiting varied phenotypes, genotypes, and functional properties[2]. Intratumoral heterogeneity complicates accurate diagnosis and prognosis, as well as significantly impacts treatment response and resistance mechanisms[3,4]. To gain deeper insights into the complex landscape of CRC, researchers have turned to emerging technologies such as single-cell RNA sequencing (scRNA-seq). By enabling transcriptomic profiling of individual cells within a tumor, scRNA-seq represents a powerful approach to interrogate the heterogeneity present within the tumor microenvironment[5]. This tool provides a comprehensive overview of the diverse cell types present in a tumor sample as well as their functional states, allowing for the identification of key drivers and potential therapeutic targets.

A rich mine of new potential CRC targets can be explored in the gene class of long non-coding RNAs (lncR-NAs)[6]. These non-coding RNA molecules, once thought to be transcriptional noise, have now emerged as key regulators of gene expression and cellular processes[7]. Dysregulation of lncRNAs has been implicated in many cancers including CRC, affecting various pathophysiological processes such as tumor initiation, progression, and metastasis[8]. For instance, CCAT1 (colon cancer-associated transcript 1) has been identified as an oncogenic lncRNA that promotes CRC growth and metastasis through multiple mechanisms, including chromosomal looping and gene regulation[9]. BANCR (BRAF-activated non-coding RNA) is another lncRNA that has been implicated in CRC progression and metastasis, acting as a sponge for microRNAs and modulating the expression of target genes[10,11]. Due to their cell type-specific expression and function[6,12], LncRNAs present promising

[1]Department for BioMedical Research, University of Bern, 3008 Bern, Switzerland. [2]Department of Medical Oncology, Inselspital, Bern University Hospital, University of Bern, 3010 Bern, Switzerland. [3]Urology Research Laboratory, Department for BioMedical Research, University of Bern, 3008 Bern, Switzerland. [4]Department of Urology, Inselspital, Bern University Hospital, University of Bern, 3010 Bern, Switzerland. [5]Institute of Genetics and Biophysics Adriano Buzzati-Traverso (IGB-ABT), National Research Council (CNR), Naples, Italy. [6]These authors contributed equally: Panagiotis Chouvardas and Roberta Esposito. ✉email: panagiotis.chouvardas@unibe.ch; roberta.esposito@unibe.ch

candidates for targeted therapies and precision medicine approaches. To date, little is known about lncRNA expression in the intra-tumor heterogeneity.

Here, through integrating cutting-edge approaches, we delve into the intricate relationship between tumor heterogeneity and lncRNAs in CRC. We analyzed available scRNA-seq data, providing the panorama of epithelial lncRNAs. Up-regulated epithelial lncRNAs in CRC compared to normal tissues were identified by using bulk RNA-sequencing CRC profiles. Ultimately, we used CRISPRi technology to validate the involvement of two lncRNAs, *CASC19* and *LINC00460*, in CRC disease progression, thereby demonstrating the potential of our approach in the identification of functional lncRNAs and novel RNA drug targets.

Results

LncRNAs are a more specific cell type marker compared to protein coding genes

Aiming to explore the role of lncRNAs in CRC heterogeneity, we analyzed single-cell RNA sequencing (scRNA-seq) data from 29 samples, derived from tumor, border regions or matched normal mucosa. After quality control we analyzed 49′436 single-cell transcriptomes, measuring 28′751 genes (17′512 protein-coding, 7797 lncRNAs and 3442 other biotypes) . The single-cell profiles are of high quality as is reflected by the number of features and low mitochondrial content (Suppl. Figure 1A). Uniform Manifold Approximation and Projection (UMAP) analysis shows that the cells cluster by cell type (Fig. 1A), as defined in the original publication. No batch effect was observed among the two datasets after merging them (Suppl. Figure 1B). The cells were re-annotated by an automated method and the cell type labels were checked for consistency between the two approaches (Suppl. Figure 1C). Hence, we concluded that this is a high-quality dataset, with no batch effect and with accurate cell type annotation labels. We then sought to explore the cell type specific expression of protein-coding and lncRNA genes by estimating the Gini coefficient of their expression in the different cellular compartments. LncRNAs show
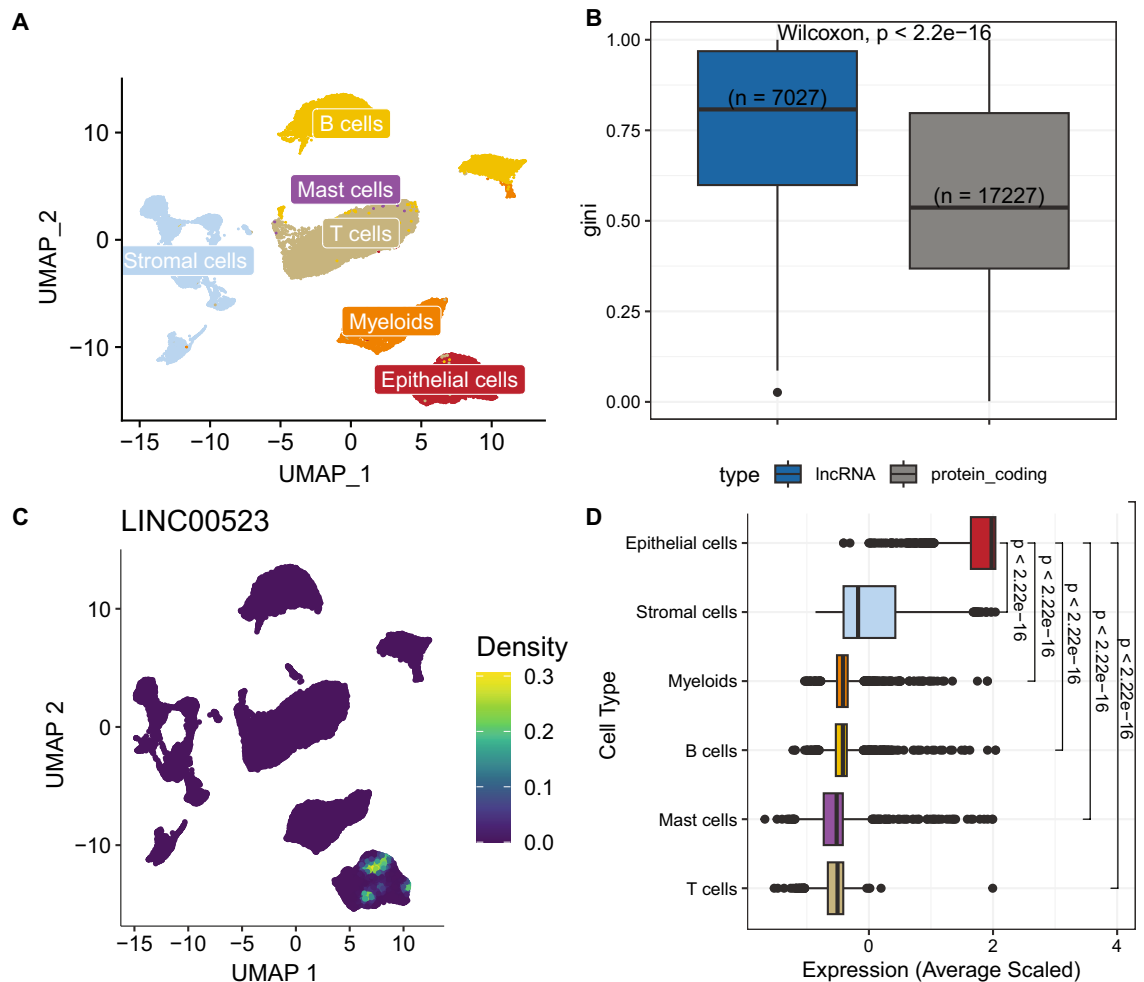

Figure 1. ScRNA-seq and cell type enrichment analysis. (**A**) UMAP of the single cells colored by cell type. (**B**) Comparing cell type specificity between lncRNAs and protein coding genes. (**C**) Example of a lncRNA with high Gini coefficient. (**D**) Expression overview of epithelial lncRNAs across cell types. Expression is averaged within the cell types and scaled across the whole dataset. Statistical tests (**B**, **D**) are performed using two-sided Wilcoxon tests.

a strikingly higher Gini coefficient of expression compared to protein-coding genes (Fig. 1B). Our results suggest that lncRNAs are overall expressed in a much more cell type specific manner compared to protein-coding genes.

As an example, LINC00523 has a Gini coefficient of 1, suggesting that it is solely expressed in only one cell type. Indeed, when estimating the density of expression of LINC00523, it is only found in epithelial cells (Fig. 1C). Differential expression analysis was then performed to identify lncRNAs enriched in epithelial cells. We identified 996 lncRNAs that show a clear overexpression in epithelial cells while they are only mildly expressed in the other cell types (Fig. 1D). Moreover, we further explored the expression of the epithelial lncRNAs in the different sample classes (i.e. tumor, border and normal) and they show on average a significantly higher expression in tumor epithelial cells (Suppl. Figure 1D). Taken together, our approach generated a unique list of putative RNA targets enriched in CRC epithelium.

Integration of scRNA-seq data with large-scale transcriptomics datasets reveals disease relevant lncRNAs

To further prioritize the list of epithelial lncRNAs and identify disease relevant genes, we accessed differential gene expression results from The Cancer Genome Atlas (TCGA) consortium, based on 270 bulk RNA-sequencing profiles, utilizing the large number of samples and therefore greater statistical power. We identified 426 differentially expressed lncRNAs, from which 303 are down- and 126 are up-regulated in the tumor samples. When we overlap these genes with the identified 996 epithelial lncRNAs, we observe 98 common genes (77 up- and 21 down-regulated) (Fig. 2A,B). We hypothesize that promising drug targets should show a minimal expression in normal tissue to further reduce the probability of unintended side effects. Moreover, we focus on the upregulated genes in the tumor samples because it is much more feasible to design therapeutic strategies which inhibit than increasing the expression of a lncRNA.

Therefore, from the 77 epithelial lncRNAs which were upregulated in tumor samples we prioritized 11 candidates: *AC016735.1*, *AC123023.1*, *BBOX1−AS1*, *CASC19*, *LINC00659*, *FEZF1−AS1*, *LINC00460*, *RP11−595B24.2*, *RP11−796E10.1*, *RP5−940J5.9* and *RP11−350J20.12* (Fig. 2C). Taken together, these lncRNAs are enriched in the
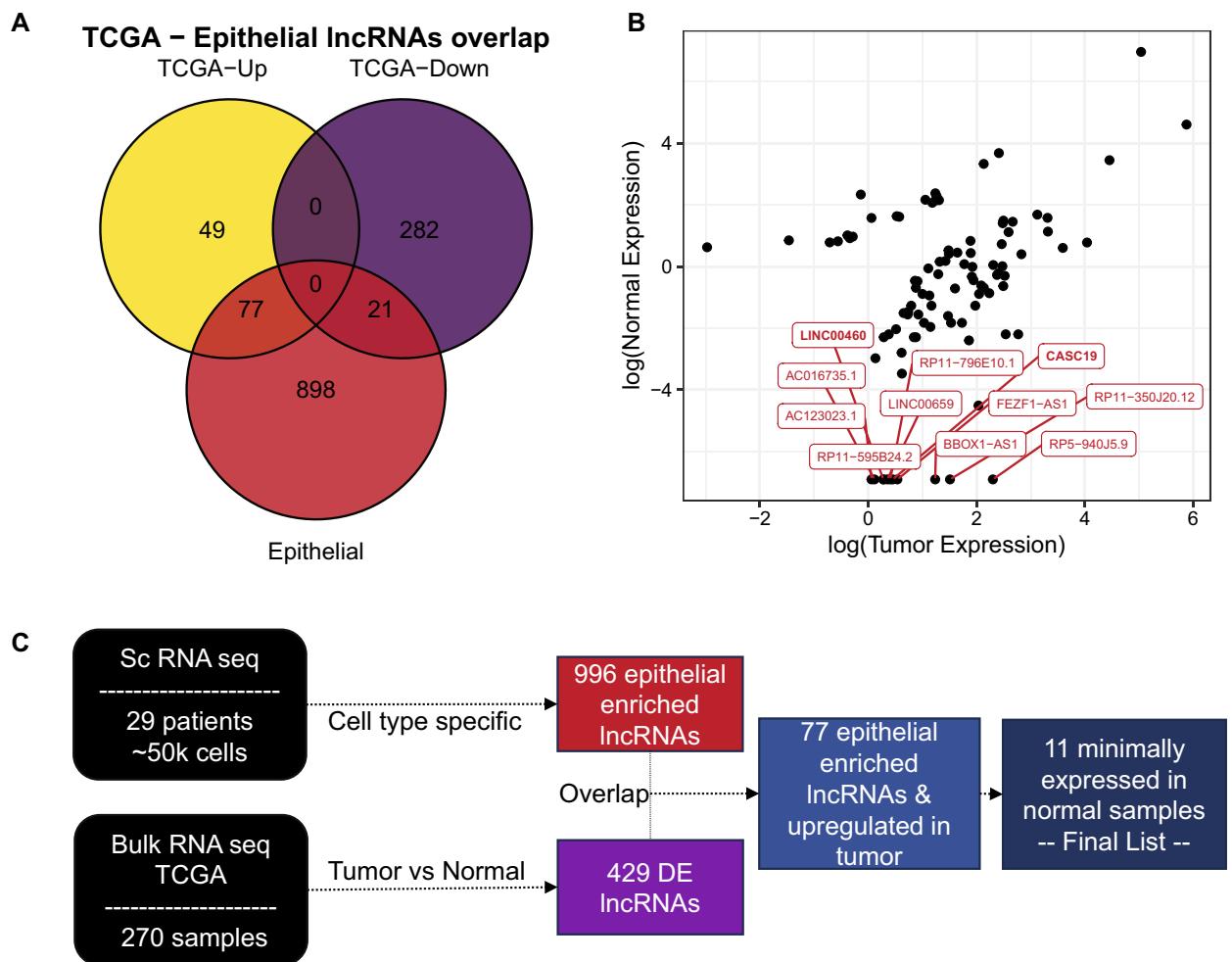

Figure 2. Integration with TCGA and prioritization of lncRNAs. (**A**) Overlap of epithelial (red) and differentially expressed lncRNAs in TCGA (yellow/purple). (**B**) Tumor vs normal expression from TCGA, highlighting 11 lncRNAs which show minimal expression in normal tissues. Expression values are normalized as transcripts per million and log-transformed. (**C**) Schematic of final candidates' selection.

epithelial cells of CRC patients, upregulated in the tumor and minimally expressed in normal tissue. Therefore, they represent highly promising candidates as novel CRC RNA drug targets.

Validating the expression of prioritized lncRNAs in CRC cell lines and its efficient reduction through CRISPR knockdown

To ensure the reliability of our results, we selected two lncRNAs from the 11 that show a specific expression in CRC samples (Fig. 2C) for further validation: *CASC19* (Cancer Susceptibility candidate 19; ENSG00000254166.2) and *LINC00460* (long intergenic non-protein coding RNA 460; ENSG00000233532.5). We chose *CASC19* and *LINC00460* based on several criteria: (i) their high expression levels in CRC tissues and cell lines; and (ii) the feasibility of designing CRISPRi sgRNAs targeting these genes without impacting the expression of neighbouring genes (i.e. transcriptional start site, TSS, > 10 Kb from the nearest protein coding gene).

CASC19 is a lncRNA that is located on chromosome 8q24.2, containing at least 29 different annotated transcripts in Gencode v38 annotation[13] (Fig. 3A). *CASC19* has been described as an oncogenic gene involved in CRC progression[14,15]. In normal tissues, *CASC19* is expressed mainly in esophagus (Suppl. Fig. 2A), while it is upregulated in cancerous tissues, especially in CRC (Suppl. Fig. 2B).

LINC00460 is an intergenic lncRNA, located on chromosome 13q33.2, which has been previously described in human malignancies[16]. It contains two annotated transcripts, partially sharing the first exon: a shorter one containing three exons made up of 1′746 nucleotides and a longer one of seven exons made up of 2′124 nucleotides base pairs, mainly expressed in normal brain tissue (Suppl. Fig. 2C). They both show a significant upregulation in CRC tissues when compared to healthy tissues (Suppl. Figure 2D).

We assessed the expression levels of *CASC19* and *LINC00460* genes in a pool of colon cell lines: DLD-1, COLO-205, HCT116, and LoVo as colon cancer models and CCD841 as a normal epithelial cell line. Both genes are upregulated in CRC cell lines compared to the normal epithelial colon cell line CCD841 with the expression being the most pronounced in DLD-1 cell line (on average 114.57 fold change for *CASC19* and 7.98 fold change for *LINC00460*) (Fig. 3B)**.**

Competition assay indicates CASC19 and LINC00460 are involved in CRC proliferation

To assess the impact of these genes on cancer cell phenotype, we proceeded with a CRISPRi-mediated knockdown (Fig. 3C) in the DLD-1 and COLO-205 cell lines. We established DLD-1 and COLO-205 to stably express ZIM3-KRAB-BFP-dCas9 (enriched for BFP expression; Suppl. Figure 3A). The KRAB domain derived from the ZIM3 protein was selected as a repressor domain based on its notable capability to induce robust knockdown effects without causing non-specific impacts on cell viability and gene expression[17]. We transduced these cell lines with lentiviral vectors carrying different sgRNAs. Each gene was targeted with two different high-scoring sgRNAs, that mediated a potent downregulation of *CASC19* and *LINC00460* genes in both cell lines (Fig. 3D,E). We could observe a pronounced reduction in gene expression in both DLD-1 (Fig. 3D) and COLO-205 cell lines (Fig. 3E) that were then used to assess the impact of gene downregulation on growth phenotype (Fig. 3F, Suppl. Figure 3B).

Growth phenotypes resulting from CRISPRi-mediated knockdown of *CASC19* and *LINC00460* genes were tested by competition assay. Cells transduced with sgRNAs targeting both lncRNAs were tagged with mCherry, while cells transduced with control were tagged with GFP (Fig. 3F, Suppl. Figure 3B). mCherry + and GPF + cells are mixed in a 1:1 proportion on day 0 and the change in ratio of mCherry + to GPF + cells was followed over a one-week period with measurements taken at two-day intervals. We could show that both sgRNAs for *CASC19* and *LINC00460* yielded strong effects on cell fitness: mCherry + cells expressing sgRNAs targeting our candidates were out-competed by control cells (GFP +, targeting AAVS1 locus). The effect was comparable to inhibition of the essential ribosomal gene RPS5, highlighting the bona fide of our approach to identify valuable lncRNA targets in epithelial cancer cells.

To assess the effect of *CASC19* and *LINC00460* on cell migration, we conducted both wound healing (or scratch-wound) (Fig. 3G) and transwell assays (Fig. 3H) using the DLD-1 cell line. Our findings reveal that inhibiting *CASC19* expression significantly impedes cancer cell migration, whereas the knockdown of *LINC00460* does not produce a similar effect.

Discussion

Despite the some advantages, most scRNA-seq studies were still limited to a focus on protein-coding genes and overlook the potential role of lncRNAs. This study aimed to investigate the role of long non-coding RNAs (lncRNAs) which have emerged as central players and key regulators in multiple biological processes such as cancer cell proliferation and drug resistance[8,18] in colorectal cancer (CRC) heterogeneity.

By analyzing a large dataset of single-cell and bulk transcriptomics profiles, we demonstrated the cell type- and tissue-specificity of lncRNAs in CRC and identified the first set of lncRNAs enriched in CRC epithelial cells. Targeting such cell-type enriched lncRNAs could potentially offer novel therapeutic strategies that exploit the unique vulnerabilities of CRC epithelial cells. Moreover, we focused on the differential expression of certain lncRNAs in tumor versus normal samples and validated their functional involvement in CRC proliferation.

Our work serves as a foundation for future investigations focusing on the functional characterization of these epithelial-enriched lncRNAs and their potential implications in CRC progression. This study highlights the importance of considering lncRNAs in understanding the heterogeneity and functional diversity of CRC cells.

Notably, we spotlighted two lncRNAs, *CASC19* and *LINC00460*, upregulated in CRC cell lines and associated with aggressive forms of CRC and poorer patient prognosis [14,15]. *CASC19* is involved in enhancing the invasive and migratory potential of CRC cells through positive regulation of cell migration inducing hyaluronidase 1 (CEMIP) and epithelial-mesenchymal transition markers[19]. Similarly, *LINC00460* is described to be involved in modulating cell migration and invasion by regulating miR-613/SphK1[20] and miR-939-5p/LIMK2 axes[21].
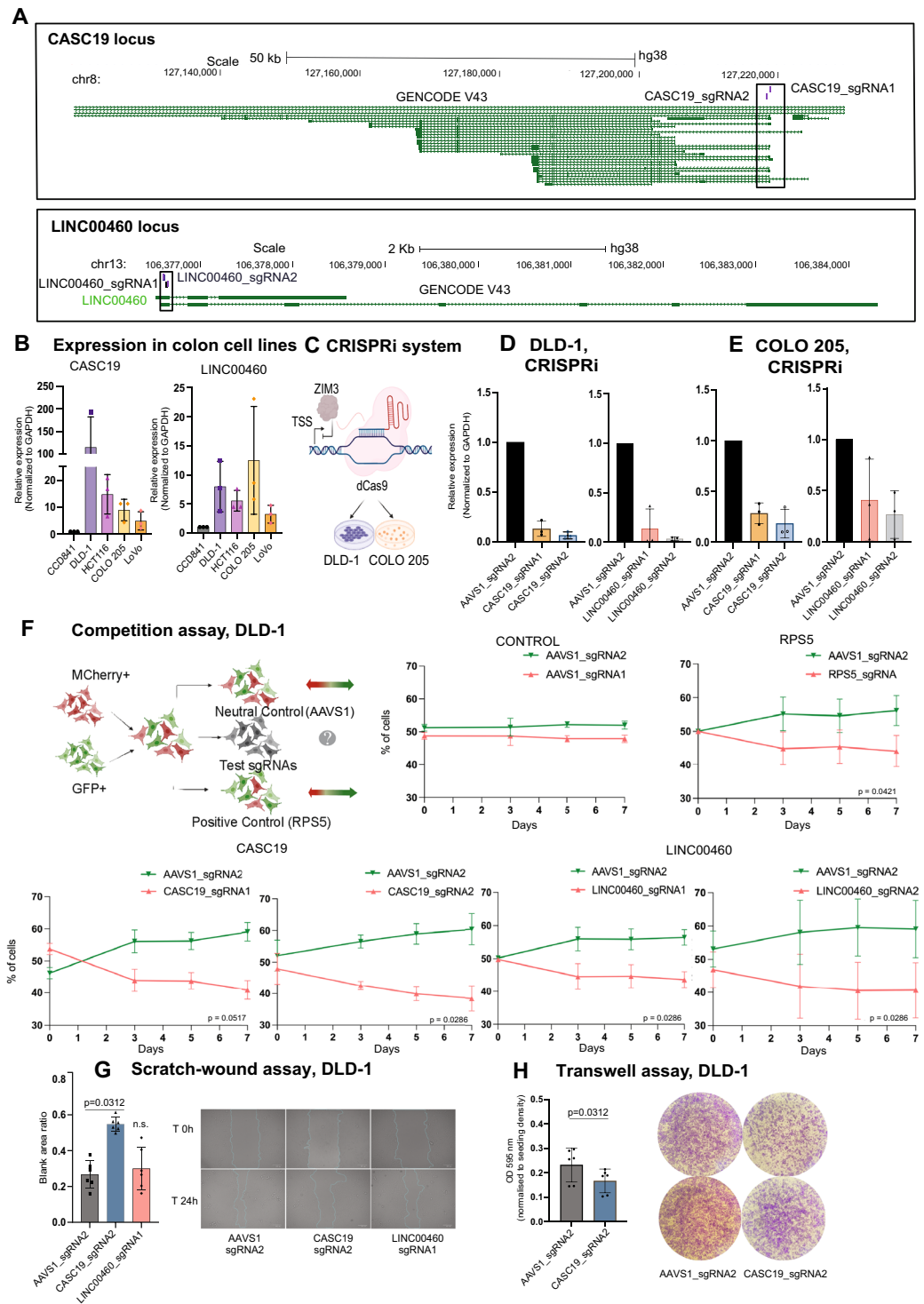
Figure 3. Expression of *CASC19* and *LINC00460* in CRC cell lines and their involvement in CRC proliferation (**A**) The genomic locus of *CASC19* and *LINC00460* lncRNA genes. The sgRNA position is highlighted by the black boxes. (**B**) The expression levels of *CASC19* and *LINC00460* were assessed by qRT-PCR in various CRC cell lines relative. The expression has been normalized to GAPDH and to the normal epithelial colon cell line CCD841. (**C**) Cartoon of CRISPRi system consisting of dCas9 enzyme fused with ZIM3-derived KRAB repressor domain. (**D**) and (**E**) Expression levels of *CASC19* and *LINC00460* genes in DLD-1 and COLO-205 cell lines after CRISPRi knockdown. Each gene was targeted using two distinct sgRNAs. (**F**) Competition assay used to assess the phenotypic implications of *CASC19* and *LINC00460* CRISPRi downregulation. Initially, cells expressing the mCherry marker and carrying various single-guide RNAs (sgRNAs) were combined with GFP-positive control cells in equal proportions. The red-to-green fluorescence ratio was measured at consecutive time intervals using flow cytometry. $N = 3$ experiments were performed, and statistical significance was estimated by linear regression model on log2 values. The mean value $-/+$ SD is plotted. Replicates were performed at different times (experimental replicates). (**G**) Scratch-wound assay was conducted using DLD-1 cell line. In the left panel, data quantification is presented as a bar plot. The right panel showcases representative images of the wound healing assay. (**H**) Transwell assay was performed on DLD-1 cells with CASC19 knockdown and control cells. Data are depicted as mean ± SD, with n = 3 for experiments (**B, D, F**) and n = 6 for experiments G and H. Statistical significance was determined using the Wilcoxon matched-pairs signed rank test.

Through rigorous experimentation, including CRISPR knockdown assays, we provided compelling evidence of their functional impact on CRC proliferation, with *CASC19* additionally being implicated in cell migration. Intriguingly, the KD of *LINC00460* had a neutral impact on cell migration. This finding contrasts with what has been previously described in the literature[27] Different factors can contribute to the discrepancy of the results: cell line specificity: cell lines can vary significantly in their behaviour and response to experimental conditions. Differences in genetic background and signalling pathways can all influence cellular responses, including migration. Therefore, the observed differences in cell migration between our model (DLD-1) and the one previously used (HCT116) used may be attributed to inherent variations in cellular characteristics; experimental methods: variations in experimental techniques used for knockdown (i.e. CRISPRi vs siRNA) may lead to differences in outcomes. Further investigations are needed to address the involvement of *LINC00460* in CRC cell migration.

These results position *CASC19* and *LINC00460* as promising candidates for targeted therapies in CRC. However, further investigation into their precise molecular mechanisms and therapeutic potential is warranted.

Moreover, we advocate for the integration of bulk and single-cell transcriptomics datasets to enhance our understanding of CRC heterogeneity. This holistic approach addresses the limitations of relying solely on one dataset type, offering a more comprehensive depiction of CRC transcriptomic profiles.

While our study offers crucial insights into lncRNA involvement in CRC heterogeneity and identifies potential therapeutic targets, certain limitations should be acknowledged. These include the challenge of low lncRNA expression levels and incomplete annotation catalogues in scRNA-seq datasets, which future research could address through advanced transcriptome assembly techniques. Additionally, broader functional validation and experimentation across a wider range of patient samples, possibly utilizing models like Patient Derived Organoids (PDOs), would strengthen the translational relevance of our findings. In summary, our study underscores the significance of lncRNAs in CRC, providing valuable insights into their cell-type specificity and functional relevance. By integrating diverse transcriptomic datasets, we offer a data-driven approach to identifying novel therapeutic targets, paving the way for precision medicine strategies in CRC treatment.

Materials and methods

Single-cell and bulk RNA sequencing data and analysis

Single-cell RNA sequencing data were downloaded from Gene Expression Omnibus, with accession numbers GSE132465 and GSE144735[23]. Raw data were pre-processed keeping those cells which express at least one gene and have less than 5% mitochondrial reads and those genes which are expressed in at least one cell. The two datasets were merged resulting in 49′436 single cells and 28′751 measured genes. Downstream single-cell analysis was performed using Seurat[24] (version 4.1.1). Cell type annotation was assessed by the original publication and validated by scMRMA[25] (version 1.0). Gene types were assessed by parsing the Gencode annotation[13] (version 38). Cell type specificity was estimated by calculating the Gini coefficient via the DescTools package (version 0.99.46). Of note, for some genes estimating the Gini coefficient was not possible because they were not expressed in any cell after the quality control filtering. TCGA data were assessed via GEPIA2[26] web server. All analyses were performed in R (version 4.1.2). Visualizations were generated using the R packages ggplot2 (version 3.4.0), pals (version 1.7) and ggvenn (version 0.1.9). The code for the analysis and generation of the figures can be found in github https://github.com/pchouvardas/COAD_lncRNAs.

Cell line maintenance

DLD-1 and COLO-205 were purchased at ATCC and regularly tested for mycoplasma infection. Cells were grown in RPMI-1640 Medium (Sigma) supplemented with 10% (v/v) Fetal Bovine Serum (Gibco), 2 mM L-glutamine (Thermo Fisher), 100 Units/mL penicillin (Thermo Fisher) and 100 mg/mL streptomycin (Thermo Fisher). HCT116 cells were grown in McCoy's 5A Medium (Sigma) supplemented with 10% (v/v) Fetal Bovine Serum (Gibco), 2 mM L-glutamine (Thermo Fisher), 100 Units/mL penicillin (Thermo Fisher) and 100 mg/mL streptomycin (Thermo Fisher). LoVo cells were grown in Kaighn's Modification of Ham's F-12 Medium (Sigma) supplemented with 10% (v/v) Fetal Bovine Serum (Gibco), 2 mM L-glutamine (Thermo Fisher), 100 Units/mL penicillin (Thermo Fisher) and 100 mg/mL streptomycin (Thermo Fisher). The cells were grown at 37 °C in the presence of 5% CO_2. HEK293T cells and CCD841 cells were grown in Dulbecco's Modified Eagle's Medium (Sigma) supplemented with 10% (v/v) Fetal Bovine Serum (Gibco) and 2 mM L-glutamine (Thermo Fisher) 100 Units/mL penicillin (Thermo Fisher) and 100 mg/mL streptomycin (Thermo Fisher. All the cells were grown at 37 °C in the presence of 5% CO_2.

Lentivirus production

24 h before transfection 2.5×10^6 HEK293T cells were seeded 100 mm tissue culture-treated plates coated with 0.1 mg/L poly-D-lysine (Gibco). Cells were triple-transfected with 12.5 µg of the plasmid of interest (UCOE-SFFV-ZIM3-KRAB-dCas9-P2A-BFP plasmid or pDECKO_mCherry plasmid containing sgRNA), 4 µg of pCMV-VSV-G (Addgene #8454) and 7.5 µg of psPAX2 (Addgene #12260) which were mixed with Lipofectamine 2000 (Thermo Fisher) and Opti-MEM™ I Reduced-Serum Medium (Gibco) according to the manufacturer instructions. 4–6 h after transfection, the medium was replaced with complete DMEM (Gibco). Viral supernatant was collected 24-, 48- and 72 h post-transfection. Collected viral supernatants were eventually pooled together and centrifuged at 3000 rpm for 15 min. The supernatant was collected and mixed with cold PEG-it Virus Precipitation Solution (4x). The mix was refrigerated overnight at 4 °C and centrifuged at 1500 g for 30 min at 4 °C. The supernatant was removed, followed by another round of centrifugation at 1500 g for 5 min. Any remaining supernatant was removed, and the lentiviral pellet was suspended in 200µL of cold $1 \times$ PBS (Gibco), aliquoted into cryogenic vials and stored at -80 °C.

Generation of cell lines stably expressing dCas9

DLD-1 and COLO-205 cell lines stably expressing dCas9 fused to ZIM3 repressor domain, were generated through transduction of parental DLD-1 and COLO-205 cells (1.5×10^6) with a concentrated lentiviral vector (30 μL) carrying UCOE-SFFV-Zim3-dCas9-P2A-BFP plasmid suspended in 10 mL of complete RPMI-1640 medium (Gibco) containing 8 μg/mL Polybrene. BFP-positive cells were selected through several rounds of expansion and FACS sorting.

Design and cloning of individual sgRNAs

sgRNAs for CRISPRi were designed using the CRISPick tool from Broad Institute (https://portals.broadinstitute.org/gppx/crispick/public). For each sgRNA, forward and reverse DNA oligos were synthesised with 20 bp over-laps with the 3' end of the U6 promoter and the 5' end of the sgRNA scaffold sequence. For cloning of annealed oligos into the pDecko_mCherry backbone (Addgene #78534), the plasmid was first digested with BsmBI, loaded on a 0.8% agarose gel and purified. 100 ng of plasmid was combined with 1 mL of each DNA oligo (1 mM), and 10 mL of homemade Gibson Assembly Mastermix (2x). The cloning product was checked by Sanger sequencing. Validated plasmid constructs were transformed into competent Stbl3™ *E. coli* cells—prepared with Mix & Go! *E. coli* Transformation Kit (Zymo Research). Plasmids were prepared following standard protocols. Generated plasmids were then used for lentivirus production.

sgRNA sequences are listed in Supplementary Table 1.

Knockdown of lncRNA genes in CRC cell lines stably expressing dCas9

DLD-1 and COLO-205 cell lines stably expressing sgRNA targeting LINC00460 and CASC19, lncRNA genes were generated through transduction of parental DLD-1 and COLO-205 cells (0.5×10^6) with a concentrated lentiviral vector (10 μL) carrying pDecko_mCherry suspended in 2 mL of complete RPMI-1640 medium (Gibco) containing 8 μg/mL Polybrene. 24 h post-transduction, the antibiotic selection was induced by supplementing the culturing medium with 2 mg/mL puromycin (Thermo Fisher) for at least 3 days.

RT-qPCR gene expression analysis

Total RNA was extracted from 500′000 cells of each cell line with Quick-RNA™ Miniprep Kit (Zymo Research). Isolated RNA was then retro-transcribed into cDNA with GoScript™ Reverse Transcription System (Promega) and the expression of the CRISPRi targeted lncRNA genes was assessed through Real-Time PCR with the GoTaq® qPCR Master Mix (Promega). Results were presented as the mean fold change compared to controls utilizing the $\Delta\Delta$Ct method, derived from a minimum of three biological replicate samples. Glyceraldehyde- GAPDH expression served as an internal calibrator. Standard deviation (SD) was utilized to represent error bars unless specified otherwise. Primers sequences are listed in Supplementary Table 1.

Competition assay

The competition assay was carried out following the protocol previously established in the lab[27,28]. After transduction with lentiviral particles containing pDecko_mCherry sgRNA constructs targeting *LINC00460* and *CASC19* and seven days after Puromycin selection (2 μg/mL), 250 000 cells stably expressing effector and sgRNA were mixed with 250 000 GFP expressing cells. Cell growth was evaluated by measuring the ratio of GFP- and mCherry-positive cells over seven days by flow cytometry on LSR II SORP instrument (BD Biosciences). As a positive control, cells were established to stably express the CRISPRi effector in conjunction with a single guide RNA (sgRNA) that specifically targeted the essential ribosomal gene RPS5. Conversely, as a negative control, cells were also established to stably express the CRISPRi effector along with an sgRNA that targeted the AAVS1 locus. $N = 3$ experiments were performed, and statistical significance was estimated by linear regression model on log2 values. The mean value $-/+$ SD is plotted. Replicates were performed at different times (experimental replicates).

Scratch-wound assay

A total of 3.0×10^6 DLD-1 cells were seeded into a 6-well culture plate and cultivated for 24 h to attain a 100% monolayer confluency. Cell proliferation was inhibited with Actinomycin D (1 μg/mL). Subsequently, a gentle cross-scratching procedure was carried out using a 200-μl tip. Following this, the plates were washed with PBS, supplemented with RPMI 1640 Media, and microscopic images were captured using an inverted light microscope at both 0 h and 24 h after the scratching process. The blank area of the scratch was calculated using ImageJ soft-ware and the ratio between two time points was determined. Statistical significance was assessed by Wilcoxon matched-pairs signed rank test.

Transwell assay

The cell migration assay was performed utilizing Transwell inserts with an 8 μm pore filter (Corning Costar). DLD-1 cells cultured in complete RPMI 1640 media underwent serum deprivation 12 h prior to the experiment. Meanwhile, the transwells were pre-incubated overnight in 600 μl of complete RPMI 1640 media in a 24-well plate. Subsequently, 100,000 cells were seeded in 100 μl of serum-free RPMI 1640 media onto the upper chamber of the Transwell insert. The insert was then immersed in 600 μl of complete RPMI 1640 media (10% FBS) within a 24-well plate and maintained at 37 °C for 36 h. The seeding density was validated using the CellTiter-Glo Luminescent cell viability assay (Promega).

After 36 h, the upper chamber of the Transwell was washed twice with 300 μl of PBS. Subsequently, cells were stained with 1% crystal violet for 30 min. Cells on the upper chamber were carefully removed using a cotton swab, and the Transwells were air dried. Images of cells in the lower chamber were captured using an inverted

microscope. The cells stained with crystal violet were then destained by incubating Transwells in 350 µl of PBS with 1% SDS for 30 min. Following destaining, absorbance was measured at 595 nm, and the values were normalized to those obtained from the cell titer assay. Statistical significance was evaluated using the Wilcoxon matched-pairs signed rank test.

Data availability

Single-cell RNA sequencing data were downloaded from Gene Expression Omnibus, with accession numbers GSE132465 and GSE144735[23]. TCGA data were assessed via GEPIA2[26] web server. All the data generated in this study can be found https://github.com/pchouvardas/COAD_lncRNAs.

Code availability

The code for the analysis and generation of the figures can be found in github https://github.com/pchouvardas/COAD_lncRNAs.

References

1. Colorectal cancer statistics. *WCRF International* https://www.wcrf.org/cancer-trends/colorectal-cancer-statistics/.
2. Sobral, D. *et al.* Genetic and microenvironmental intra-tumor heterogeneity impacts colorectal cancer evolution and metastatic development. *Commun. Biol.* **5**, 937 (2022).
3. Burrell, R. A., McGranahan, N., Bartek, J. & Swanton, C. The causes and consequences of genetic heterogeneity in cancer evolution. *Nature* **501**, 338–345 (2013).
4. Meacham, C. E. & Morrison, S. J. Tumour heterogeneity and cancer cell plasticity. *Nature* **501**, 328–337 (2013).
5. Puram, S. V. *et al.* Single-cell transcriptomic analysis of primary and metastatic tumor ecosystems in head and neck cancer. *Cell* **171**, 1611-1624.e24 (2017).
6. Derrien, T. *et al.* The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression. *Genome. Res.* **22**, 1775–1789 (2012).
7. Statello, L., Guo, C.-J., Chen, L.-L. & Huarte, M. Gene regulation by long non-coding RNAs and its biological functions. *Nat. Rev. Mol. Cell Biol.* **22**, 96–118 (2021).
8. Gutschner, T. & Diederichs, S. The hallmarks of cancer: A long non-coding RNA point of view. *RNA Biol.* **9**, 703–719 (2012).
9. Xiang, J.-F. *et al.* Human colorectal cancer-specific CCAT1-L lncRNA regulates long-range chromatin interactions at the MYC locus. *Cell Res.* **24**, 513–531 (2014).
10. Flockhart, R. J. *et al.* BRAFV600E remodels the melanocyte transcriptome and induces BANCR to regulate melanoma cell migration. *Genome. Res.* **22**, 1006–1014 (2012).
11. McCarthy, N. Epigenetics. Going places with BANCR. *Nat. Rev. Cancer* **12**, 451 (2012).
12. Liu, S. J. & Lim, D. A. Modulating the expression of long non-coding RNAs for functional studies. *EMBO Rep.* **19**, e46955 (2018).
13. Frankish, A. *et al.* GENCODE 2021. *Nucleic Acids Res.* **49**, D916–D923 (2021).
14. Wang, J. J. *et al.* Expression and function of long non-coding RNA CASC19 in colorectal cancer. *Zhongguo Yi Xue Ke Xue Yuan Xue Bao* **39**, 756–761 (2017).
15. Zhao, J., Li, C., Chen, S., Xu, Y. & Liu, F. Overexpression of CASC19 indicates poor prognosis and facilitates proliferation, migration and invasion in colorectal cancer. *Transl Cancer Res.* **8**, 1249–1257 (2019).
16. Su, M. *et al.* Oncogenic roles of the lncRNA LINC00460 in human cancers. *Cancer Cell Int.* **22**, 240 (2022).
17. Replogle, J. M. *et al.* Maximizing CRISPRi efficacy and accessibility with dual-sgRNA libraries and optimal effectors. *eLife* **11**, e81856 (2022).
18. Huarte, M. The emerging role of lncRNAs in cancer. *Nat. Med.* **21**, 1253–1261 (2015).
19. Wang, X.-D., Lu, J., Lin, Y.-S., Gao, C. & Qi, F. Functional role of long non-coding RNA CASC19/miR-140-5p/CEMIP axis in colorectal cancer progression in vitro. *World J. Gastroenterol.* **25**, 1697–1714 (2019).
20. Wang, L., Chen, X., Sun, X. & Suo, J. Long noncoding RNA LINC00460 facilitates colorectal cancer progression by negatively regulating miR-613. *Onco. Targets Ther.* **13**, 7555–7569 (2020).
21. Zhang, Y., Liu, X., Li, Q. & Zhang, Y. lncRNA LINC00460 promoted colorectal cancer cells metastasis via miR-939-5p sponging. *Cancer Manag. Res.* **11**, 1779–1789 (2019).
22. Aghagolzadeh, P. *et al.* Assessment of the cardiac noncoding transcriptome by single-Cell RNA sequencing identifies FIXER, a conserved profibrogenic long noncoding RNA. *Circulation* **148**, 778–797 (2023).
23. Lee, H.-O. *et al.* Lineage-dependent gene expression programs influence the immune landscape of colorectal cancer. *Nat. Genet.* **52**, 594–603 (2020).
24. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587.e29 (2021).
25. Li, J., Sheng, Q., Shyr, Y. & Liu, Q. scMRMA: single cell multiresolution marker-based annotation. *Nucleic Acids Res.* **50**, e7 (2022).
26. Tang, Z., Kang, B., Li, C., Chen, T. & Zhang, Z. GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis. *Nucleic Acids Res.* **47**, W556–W560 (2019).
27. Hou, P. *et al.* LINC00460/DHX9/IGF2BP2 complex promotes colorectal cancer proliferation and metastasis by mediating HMGA1 mRNA stability depending on m6A modification. *J. Exp. Clin. Cancer Res.* https://doi.org/10.1186/S13046-021-01857-2 (2021).
28. Esposito, R. *et al.* Multi-hallmark long noncoding RNA maps reveal non-small cell lung cancer vulnerabilities. *Cell Genom.* https://doi.org/10.1101/2021.10.19.464956 (2022).
29. Esposito, R. *et al.* Tumour mutations in long noncoding RNAs enhance cell fitness. *Nat. Commun.* **14**(1), 1–21. https://doi.org/10.1038/s41467-023-39160-7 (2023).

Author contributions

A.M. and P.D.R. performed the wet lab experiments. P.C. performed bioinformatic analysis. A.M. and R.E. designed and monitored experiments. All authors analyzed and interpreted data. P.C. and R.E. wrote the manuscript and all authors contributed to the review and editing of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-61430-7.

Correspondence and requests for materials should be addressed to P.C. or R.E.

Reprints and permissions information is available at www.nature.com/reprints.

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