

# HRas Signal Transduction Promotes Hepatitis C Virus Cell Entry by Triggering Assembly of the Host Tetraspanin Receptor Complex

Laetitia Zona,<sup>1,2,12</sup> Joachim Lupberger,<sup>1,2,12</sup> Nazha Sidahmed-Adrar,<sup>3,4,13</sup> Christine Thumann,<sup>1,2,13</sup> Helen J. Harris,<sup>5</sup> Amy Barnes,<sup>5</sup> Jonathan Florentin,<sup>6</sup> Rajiv G. Tawar,<sup>1,2</sup> Fei Xiao,<sup>1,2</sup> Marine Turek,<sup>1,2</sup> Sarah C. Durand,<sup>1,2</sup> François H.T. Duong,<sup>1,2,7</sup> Markus H. Heim,<sup>7</sup> François-Loïc Cosset,<sup>8</sup> Ivan Hirsch,<sup>6</sup> Didier Samuel,<sup>3,4</sup> Laurent Brino,<sup>9</sup> Mirjam B. Zeisel,<sup>1,2</sup> François Le Naour,<sup>3,4</sup> Jane A. McKeating,<sup>5</sup> and Thomas F. Baumert<sup>1,2,10,11,\*</sup>

<sup>1</sup>Inserm, U1110, Institut de Virologie, 67000 Strasbourg, France

<sup>2</sup>Université de Strasbourg, 67000 Strasbourg, France

<sup>3</sup>Inserm, U785, 94807 Villejuif, France

<sup>4</sup>Université Paris-Sud, UMRS785, 94807 Villejuif, France

<sup>5</sup>School of Immunity and Infection, NIHR Liver Biomedical Research Unit, University of Birmingham, Birmingham B15 2TT, UK

<sup>6</sup>Inserm, U1068, CNRS UMR7258, Centre de Recherche en Cancérologie de Marseille, Institut Paoli-Calmettes, Aix-Marseille Université, 13273 Marseille, France

<sup>7</sup>Department of Biomedicine, Hepatology Laboratory, University of Basel, 4031 Basel, Switzerland

<sup>8</sup>Inserm, U758, Ecole Normale Supérieure, 69007 Lyon, France

<sup>9</sup>High Throughput Screening Platform, Institut de Génétique et de Biologie Moléculaire et Cellulaire, Inserm, U964, CNRS UMR1704, Université de Strasbourg, 67404 Illkirch, France

<sup>10</sup>Pôle Hépato-Digestif, Nouvel Hôpital Civil, 67000 Strasbourg, France

<sup>11</sup>Institut Hospitalo-Universitaire, 67000 Strasbourg, France

<sup>12</sup>These authors contributed equally to this work

<sup>13</sup>These authors contributed equally to this work

\*Correspondence: [thomas.baumert@unistra.fr](mailto:thomas.baumert@unistra.fr)

<http://dx.doi.org/10.1016/j.chom.2013.02.006>

## SUMMARY

Hepatitis C virus (HCV) entry is dependent on coreceptor complex formation between the tetraspanin superfamily member CD81 and the tight junction protein claudin-1 (CLDN1) on the host cell membrane. The receptor tyrosine kinase EGFR acts as a cofactor for HCV entry by promoting CD81-CLDN1 complex formation via unknown mechanisms. We identify the GTPase HRas, activated downstream of EGFR signaling, as a key host signal transducer for EGFR-mediated HCV entry. Proteomic analysis revealed that HRas associates with tetraspanin CD81, CLDN1, and the previously unrecognized HCV entry cofactors integrin  $\beta$ 1 and Ras-related protein Rap2B in hepatocyte membranes. HRas signaling is required for lateral membrane diffusion of CD81, which enables tetraspanin receptor complex assembly. HRas was also found to be relevant for entry of other viruses, including influenza. Our data demonstrate that viruses exploit HRas signaling for cellular entry by compartmentalization of entry factors and receptor trafficking.

## INTRODUCTION

Viral entry into target cells requires the coordinated interaction of viral and host factors. Cellular kinases play a role in virus

uptake (Chakraborty et al., 2012; Mercer and Helenius, 2008; Pelkmans et al., 2005), but the underlying molecular mechanisms and signaling pathways are only poorly understood.

Hepatitis C virus (HCV) is a major cause of liver cirrhosis and hepatocellular carcinoma (HCC) (El-Serag, 2012). Major challenges include the absence of a preventive vaccine and resistance to antiviral treatment in a large fraction of patients (Zeisel et al., 2011). HCV is an enveloped, positive-sense single-stranded RNA virus of the *Flaviviridae* family (Murray and Rice, 2011). Virus entry into hepatocytes is a multistep process that is regulated by receptor tyrosine kinases (RTKs) (Lupberger et al., 2011). Host cell factors for the initiation of infection include heparan sulfate (Barth et al., 2006), CD81 (Pileri et al., 1998), scavenger receptor type B class I (SR-BI) (Scarselli et al., 2002), claudin-1 (CLDN1) (Evans et al., 2007), occludin (OCLN) (Liu et al., 2009; Ploss et al., 2009), and Niemann-Pick C1-like 1 cholesterol absorption receptor (NPC1L1) (Sainz et al., 2012). Virus entry is a promising target for antiviral therapy since host cell receptors exhibit a high genetic barrier to viral resistance (Zeisel et al., 2011).

CD81 belongs to the tetraspanin superfamily of integral transmembrane proteins that have been implicated in a variety of physiological and pathological processes and play a role in pathogen infection (König et al., 2010; Kremontsov et al., 2010; Silvie et al., 2003). Tetraspanins are highly organized in microdomains displaying specific and direct interactions with other tetraspanins and molecular partners. Tetraspanins are implicated in membrane protein trafficking, partitioning, and clustering in tetraspanin-enriched microdomains (TEMs) that regulate signaling pathways by membrane compartmentalization (Berditschkevich and Odintsova, 2007; Chambrion and Le Naour, 2010).

CD81 has a dynamic nature in HCV entry and its lateral diffusion regulates HCV infection (Harris et al., 2012). The physical interaction of CD81 with CLDN1 in a coreceptor complex is a prerequisite and essential step for HCV entry (Harris et al., 2010; Krieger et al., 2010). We demonstrated that RTKs, like epidermal growth factor receptor (EGFR), act as cofactors for HCV entry by promoting the formation of the CD81-CLDN1 coreceptor complexes, but the molecular mechanism is unknown (Lupberger et al., 2011). Unlike CD81, EGFR does not directly bind HCV E2, and, furthermore, EGFR activity directly correlates with HCV entry (Lupberger et al., 2011). Taken together, this highlights an essential role for RTK signaling in regulating the HCV entry process. Since EGFR supports the uptake of different viruses (Karlas et al., 2010; Lupberger et al., 2011; Pelkmans et al., 2005), it is likely that EGFR signaling plays a role in the entry of other virus families. To uncover the molecular mechanism underlying EGFR-regulated virus entry, we investigated the signaling pathway(s) and cellular transducers mediating HCV entry and investigated their impact on host receptor association and motility.

## RESULTS

### EGF Predominantly Activates Ras/MAPK Signaling in HCV Permissive Hepatic Cells

To identify the host signaling pathway of RTK-mediated HCV entry, we first studied EGFR signaling in Huh7.5.1 cells—a state-of-the-art permissive cell line for HCV infection, primary human hepatocytes (PHHs), and patient-derived liver biopsies. EGFR activation leads to phosphorylation of tyrosine residues in the intracellular domain that recruit signaling molecules to the plasma membrane that prime subsequent activation of events (Morandell et al., 2008), including mitogen-activated protein kinase (MAPK), phosphoinositide-3 kinase (PI3K), and v-Akt murine thymoma viral oncogene homolog (AKT) pathway activation (Figure 1A). Analysis of signal transduction in Huh7.5.1 cells, PHHs, and liver tissue (liver biopsies 987 and 990) with phosphokinase (Figures 1B and 1C and Figure S1A available online) and phospho-RTK arrays (Figures 1C and S1B) demonstrates that EGFR and the Ras/MAPK pathway are predominantly activated after EGF stimulation in human hepatocytes in vitro and in vivo. In contrast, activation of signaling pathways described in other cell lines and tissues (PI3K/AKT, PLC/PKC, p38/JNK, STAT3/5, Cbl, c-Src/ABL, and FAK) was less relevant in the liver-derived cells, PHHs, or liver tissue (Figures 1B, 1C, and S1) as shown by phosphorylation arrays. To further corroborate the relevance of the Ras/MAPK pathway as the primary signal transducer of EGFR in the liver, we measured the phosphorylation status of extracellular signal-regulated kinases 1 and 2 (ERK1/ERK2) and AKT by dynamic phosphoflow assay following EGFR stimulation (Firaguay and Nunès, 2009) (Figures S1C–S1F) in Huh7.5.1 and PHHs. EGF stimulation of Huh7.5.1 cells and PHHs activates Ras/MAPK signaling, while even prolonged incubation of the cells with EGF failed to activate the PI3K/AKT pathway (Figures S1D–S1F). Similar results were obtained by immunoblot of phosphoproteins in EGF-stimulated Huh7.5.1 or PHH lysates (Figure S1G) and lysates from patient-derived liver tissue (liver biopsies 956, 965, and 968) that had been stimulated with EGF ex vivo (Figure 1D). We noted

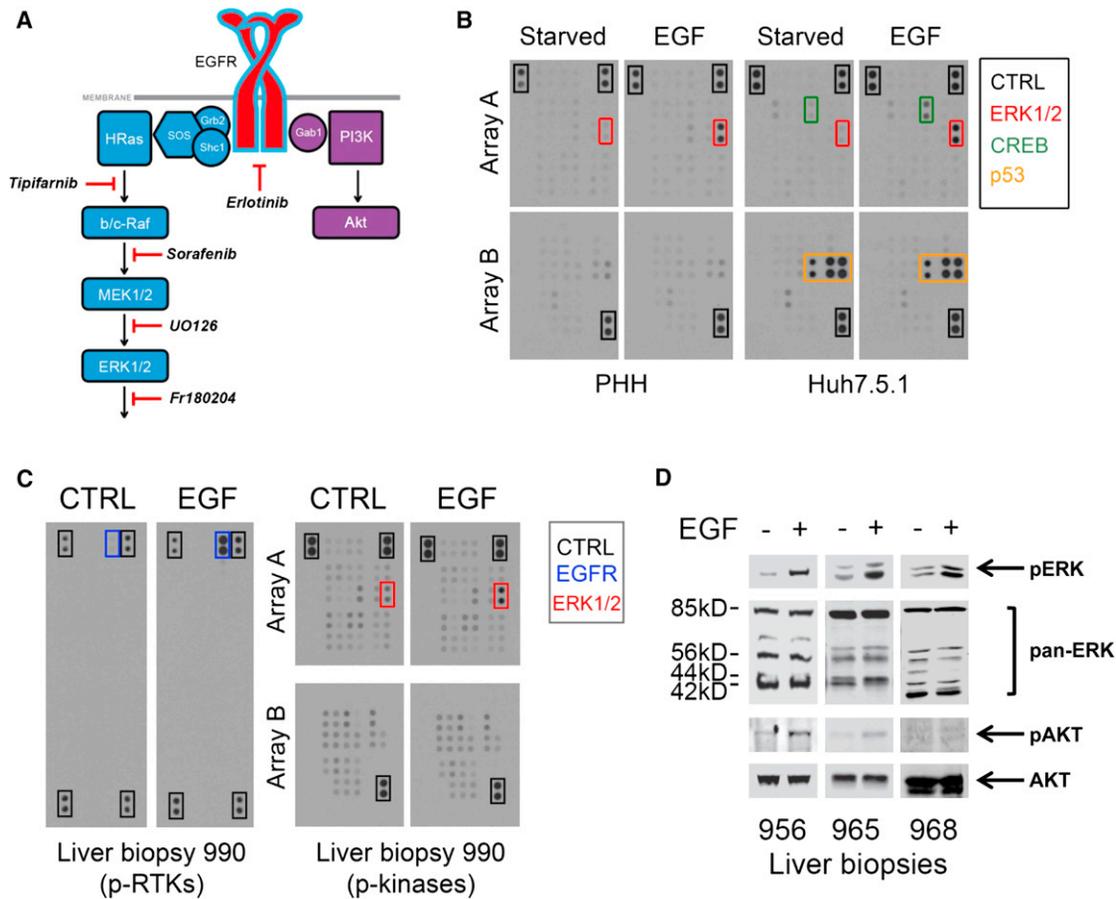
that EGF activated ERK1/ERK2 at significantly lower doses in Huh7.5.1 cells and PHHs (Figure S1G). In summary, these results demonstrate that EGF predominantly activates the Ras/MAPK pathway in hepatoma cells, PHHs, and liver tissue supporting the key relevance of the Ras/MAPK pathway for EGFR-mediated signaling in the liver.

### EGFR Scaffolding Proteins Grb2 and Shc1 Are Relevant for HCV Entry

To identify EGFR-mediated signals important for HCV entry, we performed an unbiased small interfering RNA (siRNA) screen targeting the expression of known EGFR adaptors and associated proteins (Table S1). Among the identified EGFR scaffolding proteins with impact on HCV entry, we identified growth factor receptor-bound protein 2 (Grb2) and Src homology 2 domain-containing transforming protein 1 (Shc1): known activators of Ras GTPases and MAPK pathway (Kolch, 2005) (Figure 1A). Although STAT5b scored as a binding partner with potential functional relevance (Table S1), we observed no evidence for activation of the STAT5 signaling in EGF-treated PHHs or liver tissue (Figures 1 and S1A), and a STAT5b inhibitor had no effect on HCV entry as described below. Thus, we focused on the functional relevance of Grb2 and Shc1 for HCV entry. Silencing of Grb2 or Shc1 expression significantly ( $p < 0.01$ ) decreased HCV pseudoparticle (HCVpp) entry (Figure 2A) and cell-culture-derived HCV (HCVcc) infection (Figure 2B) to similar levels as EGFR silencing, while silencing Grb2-associated binding protein 1 (Gab1) that recruits PI3K (Figure 1A) had no effect on HCV infection (Figures 2A and 2B). Specific Grb2, Shc1, and Gab1 gene silencing was validated by immunoblot (Figure 2C). In contrast, silencing of Grb2 or Shc1 expression had no effect on the entry of murine leukemia virus (MLV) pseudoparticles (Figure 2D), suggesting that the observed inhibitory effect is not related to the pseudoparticle system. Finally, we confirmed the functional relevance of Grb2 and Shc1 in HCVpp infection of polarized HepG2-CD81 cells (Figure 2E). HepG2 cells polarize in vitro and develop bile-canalculi-like spaces between adjacent cells, thus allowing the study of virus entry in a model system closely related to polarized hepatocytes in the infected liver in vivo (Mee et al., 2009). Taken together, these data show a role for EGFR scaffolding proteins Grb2 and Shc1 in HCV entry and infection.

### Inhibition of Ras and Upstream MAPK BRAf Decreases HCV Entry

Since silencing of EGFR scaffolding proteins Grb2 and Shc1 inhibits HCV entry and these proteins activate Ras/MAPK signaling, we investigated whether other members of the MAPK pathway play a role in HCV entry. We thus used a panel of well-characterized small-molecule inhibitors (Figure 1A) of EGFR (erlotinib), rat sarcoma (Ras) (tipifarnib), Raf (sorafenib), BRAf (inhibitor VI), Raf-1 (inhibitor I), mitogen-activated protein kinase 1 and 2 (MEK1/MEK2) (U0126), and ERK1/ERK2 (Fr180204) and studied their effect(s) on HCV entry and infection (Figures 3A–3F). We also ruled out a role of other major EGFR signaling pathways including PI3K/AKT, PLC/PKC, p38/JNK, STAT3/5, and FAK using well-described small-molecule inhibitors (Figure 3G). The biological activity of all used inhibitors was confirmed in functional assays in Huh7.5.1 (Figure S2).



**Figure 1. EGFR Signaling Pathways in Human HCV-Permissive Liver Cells, Hepatocytes, and Patient-Derived Liver Tissue**

(A) Scheme of the two main canonical EGFR signaling cascades: the MAPK and the PI3K/AKT pathways. Inhibitors targeting members of these pathways are indicated.

(B and C) EGFR-transduced signals in human hepatocytes and liver tissue. (B) Detection of kinase phosphorylation in Huh7.5.1 and PHHs after EGF treatment (1  $\mu$ g/ml; 15 min) with a human phosphokinase array detecting specific phosphorylation of 46 phosphorylation sites on 32 cellular kinases, which are indicated in Figure S1A. (C) Detection of RTK and kinase phosphorylation in liver tissue of patient biopsy 990 after EGF treatment (1  $\mu$ g/ml; 15 min ex vivo) with a human phospho-RTK array (detecting specific phosphorylation of 42 different RTKs as indicated in Figure S1B) and a human phosphokinase array.

(D) Analysis of the phosphorylated and total forms of ERK1/ERK2 and AKT with specific antibodies in three different liver biopsies (956, 965, and 968) after 15 min EGF stimulation (1  $\mu$ g/ml) ex vivo. Total protein (30  $\mu$ g) was separated by SDS-PAGE and stained for total and phosphorylated forms of ERK and AKT by immunoblot.

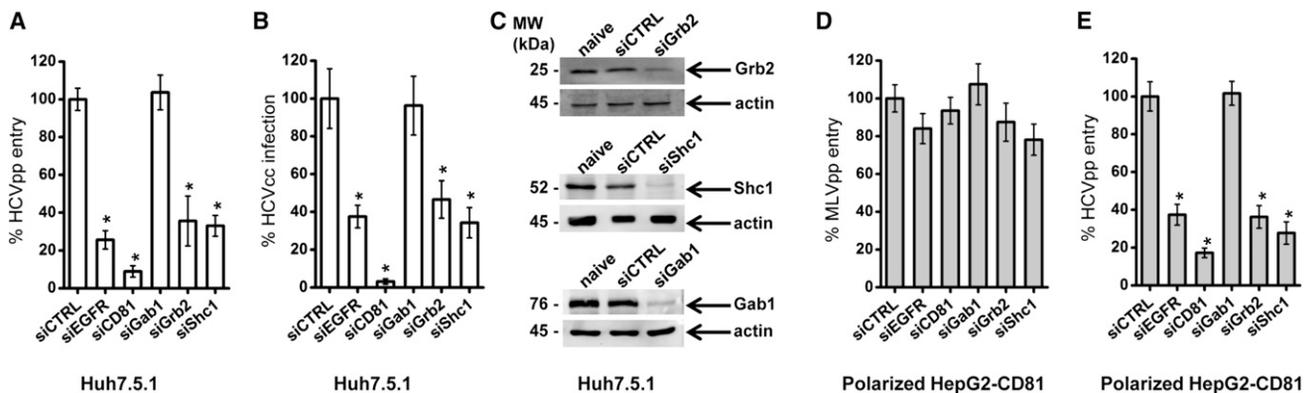
See also Figure S1 for more-detailed analyses of EGFR signaling.

All inhibitor concentrations were well tolerated in Huh7.5.1 and PHHs, as shown by MTT assays applied for inhibitors that inhibited HCV entry (Figures 3A and 3B). Since erlotinib and tipifarnib resulted in a robust, donor-independent, and highly significant ( $p < 0.0001$ ) inhibition of HCVpp entry in hepatoma cells and PHHs, we conclude that Ras is a predominant signal transducer required for EGFR-mediated HCV entry. Inhibition of Ras and upstream MAPK Raf, and to some extent MEK1/MEK2, decreased HCVpp and HCVcc infection in Huh7.5.1 cells and PHHs (Figures 3C–3E). Moreover, a different inhibition profile of MLVpp infection of PHHs suggests that the effects are not related to the lentiviral system (Figure 3F). Taken together, our observations demonstrate that the closer the inhibitor's target is toward EGFR in the MAPK signaling cascade (Figure 1A), the more efficiently HCV entry is inhibited. These data suggest that only upstream members of the MAPK signaling

pathway, but not other canonical EGFR pathways, play a major functional role in HCV entry.

#### HRas Is a Transducer of EGFR-Mediated HCV Entry

The classical Ras family comprises three isoforms: Harvey (H)-, Kirsten (K)-, and neuroblastoma (N)-Ras (Boguski and McCormick, 1993). Since Grb2, Shc1 and Raf play a role in EGFR-mediated HCV entry (Figures 2 and 3), we investigated the functional role of the Ras GTPase family in HCV entry using RNA interference (RNAi). HRas, KRas, or NRas expression (Figure S3A) was silenced in Huh7.5.1 with two individual siRNAs per target (Figures 4A–4C). The messenger RNA (mRNA) or protein expression after silencing was studied for all Ras isoforms or HRas, respectively (Figures 4A and 4B). Silencing of HRas expression markedly and significantly ( $p < 0.0005$ ) decreased HCV entry into Huh7.5.1 cells to a comparable level as EGFR



**Figure 2. EGFR Adaptors Grb2 and Shc1 Are Relevant for HCV Entry**

Silencing of EGFR adaptors Grb2 and Shc1 inhibits HCV entry. Huh7.5.1 (A–C) and polarized HepG2-CD81 (D and E) cells were transfected with individual siRNA directed against Gab1, Grb2, or Shc1 and infected with HCVpp, MLVpp, or HCVcc.

(A, C, and E) Silencing of protein expression was confirmed by immunoblot with specific antibodies targeting Gab1, Grb2, Shc1, or actin (C). HCVpp entry was assessed in Huh7.5.1 (A) and in polarized HepG2-CD81 (E) cells transfected with siRNA. siCTRL, CD81- and EGFR-specific siRNAs served as internal controls. Data are expressed as percentage HCVpp entry relative to siCTRL-transfected cells (means  $\pm$  SD from three independent experiments in triplicate,  $n = 9$ ).

(B) HCVcc infection in Huh7.5.1 cells transfected with the same siRNAs. Data are expressed as percentage HCVcc infection relative to siCTRL-transfected cells (means  $\pm$  SD from three independent experiments in triplicate,  $n = 9$ ).

(D) MLVpp entry in polarized HepG2-CD81 cells. Data are expressed as percentage MLVpp entry relative to siCTRL-transfected cells (means  $\pm$  SD from three independent experiments in triplicate,  $n = 9$ ).

\* $p < 0.01$ . See also Table S1.

silencing (Figure 4C). In contrast, silencing KRas or NRas expression had no detectable effect on HCV entry (Figure 4C). Isoform specificity of HRas gene silencing was validated by qRT-PCR (Figure S3B). Moreover, HRas silencing reduced EGF-induced ERK1/ERK2 phosphorylation, supporting a role of HRas in EGF-induced MAPK activation in hepatocytes (Figure S3C). These results demonstrate a specific role for the GTPase HRas in the HCV entry process.

Since EGFR mediates entry of viruses other than HCV, we studied the role of HRas on the entry of lentiviral pseudoparticles expressing glycoproteins from avian fowl plague virus influenza A (H7/N1), measles virus, MLV, endogenous feline virus RD114, and vesicular stomatitis virus (VSV) in Huh7.5.1 cells. Silencing HRas expression had a significant ( $p < 0.0005$ ) inhibitory effect on the entry of influenza and measles pseudoparticles (Figure 4D), suggesting that these viruses require similar signaling pathways to enter hepatoma cells as HCV. Although we previously demonstrated that EGFR silencing had no effect on measles virus entry (Lupberger et al., 2011), HRas silencing impacts measles virus entry, suggesting an EGFR-independent role of HRas in this process.

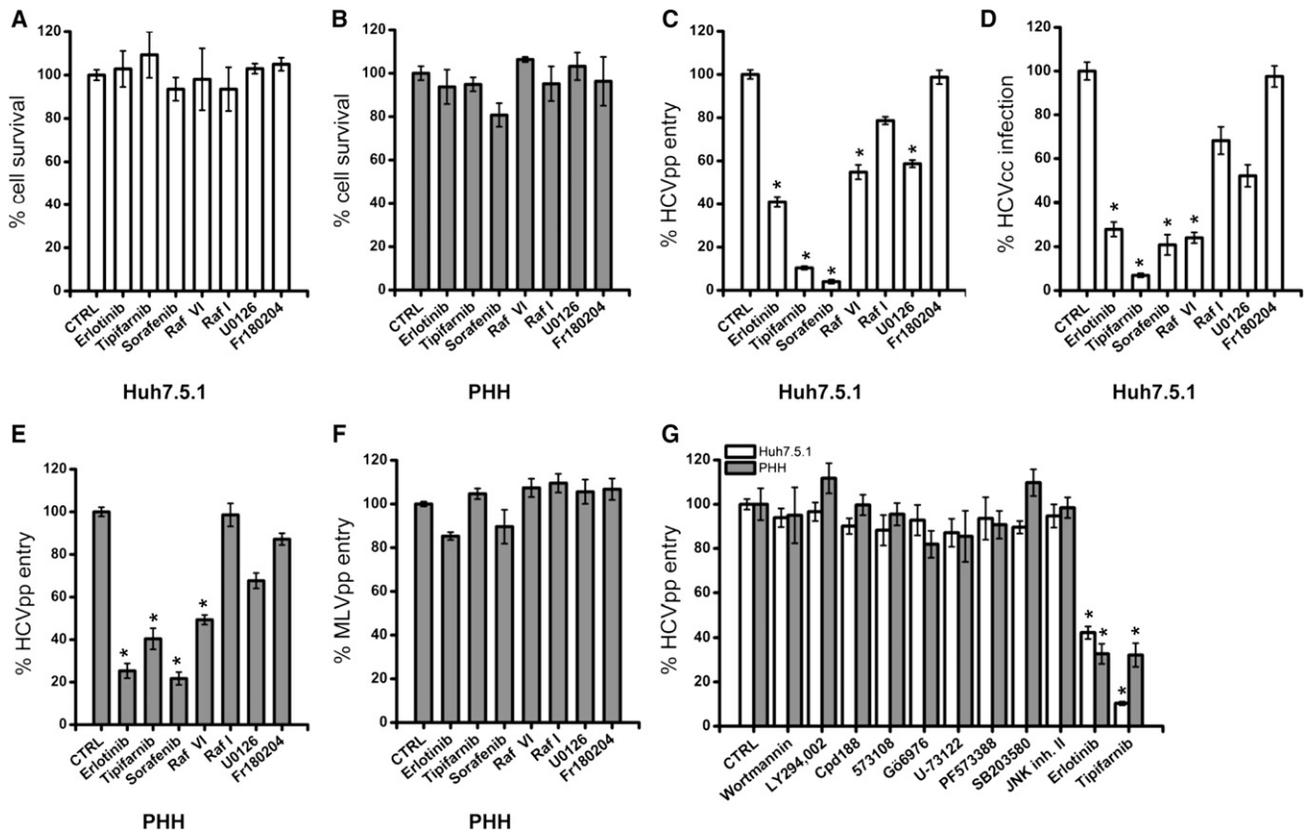
To confirm the HRas dependency of EGFR-mediated HCV entry, we performed inhibition/rescue experiments using the EGFR inhibitor erlotinib and the patient-derived transdominant active V12 mutant of HRas (Beauséjour et al., 2003). Huh7.5.1 or PHHs were transduced to express the HRas V12 mutant and 72 hr later evaluated for their HCVpp permissivity in the presence or absence of 10  $\mu$ M erlotinib (Figures 4E and 4F). Exogenous HRas expression in Huh7.5.1 and PHHs was confirmed by HRas-specific immunoblot (Figure S3D). The HRas V12 mutant increased HCVpp infection of Huh7.5.1 and PHHs in the absence of erlotinib, demonstrating an involvement of HRas in HCV entry (Figures 4E and 4F). Moreover, HRas V12

mutant rescued the inhibitory effect of erlotinib on HCV entry, confirming that HRas mediates EGFR-dependent HCV entry (Figures 4E and 4F).

Next, we investigated whether HRas mediates EGFR-dependent cell-cell transmission using a well-established cell-cell transmission assay (Lupberger et al., 2011). Reduction of (siHRas.6) or increase of (pHRas V12) HRas expression (Figures 4G and S3E–S3G) in target cells impaired significantly ( $p < 0.0005$ ) or enhanced cell-cell transmission compared to control-transduced cells, respectively. Furthermore, the HRas inhibitor tipifarnib blocked HCV cell-cell transmission (Figures 4G and S3G). These data highlight a previously unrecognized role of HRas during viral spread.

#### HRas Associates with HCV Cell Entry Factors CD81 and CLDN1

EGFR promotes the association of CD81-CLDN1 coreceptor complexes that are important for HCV entry (Lupberger et al., 2011). To investigate whether EGFR signaling modulates CD81-CLDN1 association, we analyzed whether TEMs contain members of the EGFR signaling pathways using proteomic analysis of CD81 immunoprecipitates. HepG2 and HepG2-CD81 cells were differentially labeled with stable isotope labeling with amino acids (SILAC) (Ong et al., 2002) and lysed with brij97 detergent that is reported to preserve tetraspanin interactions (Le Naour et al., 2006), and HepG2 and HepG2-CD81 lysates were pooled equally according to SILAC protocols (Ong et al., 2002). From this pool, CD81 was pulled down with beads coupled with CD81-specific IgG and coprecipitated protein complexes analyzed by mass spectrometry. Among the CD81 coprecipitated proteins were several integrins ( $\alpha$ 1,  $\alpha$ 6, and  $\beta$ 1) that are well-characterized TEM components. Therefore, we defined the threshold of specificity  $>2$  accordingly to the



**Figure 3. Upstream MAPK Are Relevant for HCV Entry in Hepatoma Cells and Human Hepatocytes**

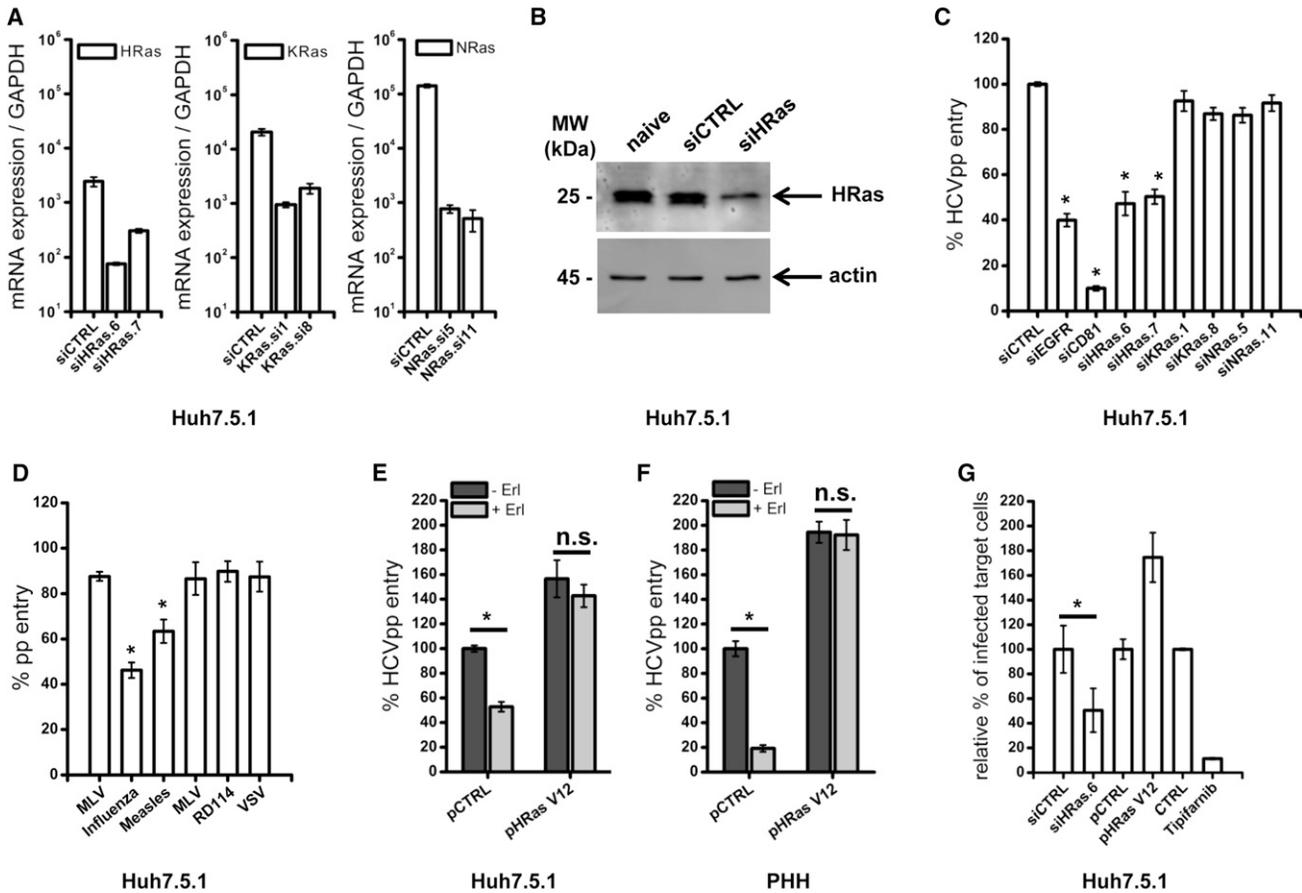
Cell survival (A and B), HCVpp entry (C, E, and G), HCVcc infection (D), or MLVpp entry (F) in Huh7.5.1 cells (A, C, D, and G) or PHHs (B, E, F, and G) incubated with small-molecule inhibitors (10  $\mu$ M) targeting EGFR (erlotinib), Ras (tipifarnib), Raf (sorafenib, 10  $\mu$ M for pseudoparticles and 1  $\mu$ M for HCVcc), BRaf (Raf inhibitor VI), Raf-1 (Raf inhibitor I), MEK1/MEK2 (U0126), or ERK1/ERK2 (Fr180204) or the major EGFR pathways, including PI3K (wortmannin and LY294,002), STAT3 (Cpd188), STAT5 (573108), PKC (Gö6976), PLC (U-73122), FAK (PF573288), p38 (SB203580), JNK (JNK inhibitor II), and Ras (tipifarnib). One hour after incubation with inhibitors, HCVpp, MLVpp, or HCVcc was added to the cells in the presence of inhibitors. Cell viability was assessed by MTT assay. EGFR (erlotinib) serves as internal control. Data are expressed as percentage HCVpp or MLVpp entry and HCVcc infection relative to solvent CTRL-treated cells (means  $\pm$  SEM from four independent experiments in triplicate, n = 12). \*p < 0.0001. See also Figure S2.

isotope ratio  $^{13}\text{C}/^{12}\text{C}$  of coprecipitated integrins and accordingly to SILAC standard procedures (Ong et al., 2002). Above this threshold, we identified tetraspanin-associated proteins such as membrane protease ADAM10, several tetraspanins (CD9, CD81, CD151), and known specific interaction partners of CD81, such as EWI-2 and CD9P-1/EWIF, validating this differential proteomic approach to identify CD81-associated proteins. Interestingly, HCV entry factors CLDN1 and SR-BI but not OCLN or NPC1L1 were identified as CD81 TEM components corroborating a close cooperation of CD81, CLDN1, and SR-BI during HCV entry (Krieger et al., 2010). These results were confirmed in a second experiment using SILAC proteomics with inverted isotope labeling (Table 1). Among the 169 components identified in TEMs, we identified HRas as the only member of the canonical EGFR signaling pathways associating with CD81. A physical interaction of HRas with the tetraspanin coreceptor complex was further supported by a partial but robust colocalization of CD81 with HRas at the plasma membrane of Huh7 cells (Figure S4A) as calculated according to Pearson (Rr = 0.25) and Manders (R = 0.659) coefficients and intensity correlation quotient (ICQ = 0.129) (Bolte and Cordelières, 2006;

Brown et al., 2010; Manders et al., 1992) (Figure S4). To further study the relevance and robustness of the HRas-CD81 colocalization, we transduced cells to express the HRas V12 mutant (Figure S4B). The colocalization of HRas V12 with CD81 was calculated (Image J software) and significantly (p < 0.005) increased as demonstrated by the Pearson (Rr = 0.544  $\pm$  0.047) and Manders (R = 0.825  $\pm$  0.056) coefficients and ICQ (0.387  $\pm$  0.067) as compared to empty vector (Rr = 0.278  $\pm$  0.054; R = 0.820  $\pm$  0.05; ICQ = 0.175  $\pm$  0.022) (Figures S4C and S4D). The distribution of HRas at the plasma membrane was observed as intense punctuated spots and correlated with high CD81 content. Taken together, these data suggest that active HRas physically associates with CD81.

**Functional Network Analyses of CD81-Associated Proteins Identified Rap2B and Integrin Beta1 as Cofactors for HCV Entry**

To identify members of TEM containing CD81 and CLDN1 with a functional role in HCV entry, we analyzed the 169 identified CD81-associated proteins (Table 1 and data not shown) for known and predicted HRas protein interactions using the



**Figure 4. HRas Is a Host Cell Factor for HCV Entry**

(A) mRNA expression of HRas, KRas and NRas compared to GAPDH after silencing of each Ras isoform with isoform-specific siRNAs. (B) Analysis of protein expression by immunoblot with specific antibodies targeting HRas or actin after Ras silencing (siHRas.6). (C) HCVpp entry in Huh7.5.1 cells transfected with individual siRNAs directed against HRas (si6 and si7), KRas (si1 and si8), and NRas (si5 and si11). siCTRL, CD81, and EGFR-specific siRNAs served as internal controls. Data are expressed as percentage HCVpp entry relative to siCTRL-transfected cells (means  $\pm$  SEM from four independent experiments in triplicate,  $n = 12$ ). (D) Cell entry of pseudoparticles expressing envelope glycoproteins of influenza, measles, MLV, RD114, and VSV in Huh7.5.1 cells transfected with an individual siRNA directed against HRas (si6). Two independent MLVpp preparations were used. siCTRL served as internal control. Data are expressed as percentage pseudoparticle entry relative to siCTRL-transfected cells (means  $\pm$  SEM from three independent experiments in triplicate,  $n = 9$ ). (E and F) Inhibition of HCV entry by erlotinib is rescued by a transdominant active HRas mutant. HCVpp entry in Huh7.5.1 cells (E) and in PHHs (F) transduced with lentiviruses expressing a transdominant active HRas mutant (pHRas V12) and treated with erlotinib (10  $\mu$ M). For HRas protein expression, see Figures S3C and S3D. Data are expressed as percentage HCVpp entry relative to pCTRL cells (means  $\pm$  SEM from four independent experiments in triplicate,  $n = 12$ ). (G) Functional role of HRas in viral cell-cell transmission. Effect of HRas silencing by siHRas.6, overexpression of HRas V12, or HRas inhibition by tipifarnib (10  $\mu$ M) on viral spread is shown. Data are expressed as percentage cell-cell transmission relative to respective controls (for RNAi, means  $\pm$  SD from three independent experiments in triplicate,  $n = 9$ ; for HRas V12 and tipifarnib, means  $\pm$  SD from one representative experiment in triplicate,  $n = 3$ ). SD for CTRL and tipifarnib are 0.77 and 0.68 respectively, and are thus not visible. \* $p < 0.0005$ . See also Figure S3.

STRING database (Jensen et al., 2009; Lupberger et al., 2011). STRING represents a metadatabase mapping all known protein-protein interactions onto a common set of genomes and proteins (Jensen et al., 2009). This analysis suggests a potential network of proteins connecting CD81 and HRas (Figure 5A) that includes known HCV entry factors CD81 and CLDN1 and HCV host factor apolipoprotein E (apoE). Functional analysis of members of this network using RNAi/HCVpp studies identified Ras-related protein (Rap2B) and integrin beta1 (ITGB1) as cofactors for HCV entry (Figure 5B). Silencing of ITGB1 and Rap2B expression was confirmed with individual siRNAs

(Figures 5C–5E). An ITGB1-specific antibody markedly and significantly ( $p < 0.0001$ ) inhibited HCV infection of Huh7.5.1 cells and PHHs (Figures 5F and 5G), validating the role of ITGB1 for HCV entry. These data suggest a functional network organized by tetraspanins in the plasma membrane consisting of CD81-CLDN1, HRas, Rap2B, and ITGB1.

**HRas Is Required for Lateral Diffusion of CD81 Promoting CD81-CLDN1 Associations**

Since CD81 plays a role in the lateral diffusion of HCV (Harris et al., 2012), we studied the effect(s) of EGFR/HRas signaling

**Table 1. Subset of CD81-Associated Proteins in HepG2-CD81 Cells Identified by SILAC Differential Proteomics, Including HRas Interacting Partners**

Protein name	ID	Sequence (%)	MW (kDa)	SILAC Labeling (Ratio H/L)	Inverted SILAC Labeling [Ratio 1 / (H/L)]
CD151	P48509	13.4	28.30	15.86	64.72
EWI-2	Q969P0	47	65.03	13.93	85.67
GNAI3	P08754	22.3	40.53	13.52	41.35
CD9	P21926	7.5	25.41	13.38	231.30
CD9P-1/ EWIF	Q9P2B2	36.70	98.55	13.16	6.72
CD81	A6NMH8	30.3	29.81	12.64	111.17
ADAM10	O14672	48.3	84.14	11.45	69.25
GNAI1	P63096	20.1	40.36	11.29	14.11
RAP2B	P61225	32.2	20.50	10.46	22.75
MPZ	Q14902	9.2	27.95	9.05	NaN
APOE	P02649	40.1	36.15	8.13	1.56
CLDN1	A5JSJ9	16.1	22.74	7.83	12.61
CD59	E9PR17	15.4	14.53	7.19	NaN
HRas	P01112	20.1	21.30	5.84	103.91
RALA	P11233	19.4	23.57	5.32	4.73
Integrin alpha6	P23229	28.8	126.63	5.09	15.78
SCAMP3	O14828	9.2	38.29	4.54	NaN
Integrin alpha1	P56199	12.4	130.85	2.98	17.16
Integrin beta1	P05556	27.7	88.41	2.61	6.51
SR-BI	Q59FM4	12.4	64.19	2.56	7.69
LMNA	P02545	19.6	74.14	2.53	0.51
SOD2	P04179	13.1	24.72	NaN	2.91
LGALS1	P09382	17.0	14.72	NaN	2.56

HepG2-CD81 and HepG2 cells were differentially labeled with stable carbon isotopes  $^{12}\text{C}$  or  $^{13}\text{C}$  (SILAC method). CD81-associated complexes were coprecipitated with CD81, digested, and analyzed by mass spectrometry. The protein ID, its molecular weight, the number of the identified peptides, and total sequence coverage for each identified protein is stated. Specificity threshold of CD81 association from each individual identified protein was defined as a peak volume ratio H/L >2 of the differentially isotope labeled versions of each protein. The results were validated by a second experiment with inverse isotope labeling (inverted). The specificity threshold for the inverted SILAC labeling was  $1 / (\text{H/L}) > 2$ . NaN, not a number. See also [Figure S4](#).

on CD81 dynamics by real-time fluorescence recovery after photobleaching (FRAP). Huh7.5.1 cells were transduced to express AcGFP-CD81, and the basal surface was imaged by TIRF microscopy. We observed a significant increase in CD81 diffusion coefficient (CTRL  $0.09 \mu\text{m}^2/\text{s}$ ; tipifarnib  $0.18 \mu\text{m}^2/\text{s}$ ,  $p < 0.05$ ) in Huh7.5.1 cells treated with the HRas inhibitor tipifarnib ([Figure 6A](#)). These data suggest that HRas signaling reduces CD81 mobility by promoting an interaction with other proteins. To investigate whether EGFR acts on CD81-CLDN1 interaction via the putative EGFR/Shc1/HRas/BRAF pathway, we studied

whether silencing of Shc1, HRas, or BRAF modulates CD81-CLDN1 coreceptor interaction using a well-established fluorescence resonance energy transfer (FRET)-based assay ([Harris et al., 2010](#)). Silencing of Shc1, HRas, or BRAF expression significantly ( $p < 0.05$ ) reduced CD81-CLDN1 FRET in Huh7.5.1 cells ([Figures 6B–6D](#), black bars) while silencing had a minimal effect on CD81-CD81 association ([Figures 6B–6D](#), open bars). These results demonstrate that HRas and BRAF play a role in the formation and/or maintenance of the CD81-CLDN1 coreceptor complexes. We previously reported that although EGFR stimulation increased HCV entry, this process does not require ligand-induced EGFR stimulation since the basal activity of the receptor, even after serum starvation, is sufficient to support HCV entry ([Lupberger et al., 2011](#)). Indeed, EGF had no significant effect on CD81-CLDN1 FRET (data not shown), supporting the hypothesis that the steady-state complex of CD81-CLDN1 is dependent on low-level constitutive EGFR signaling.

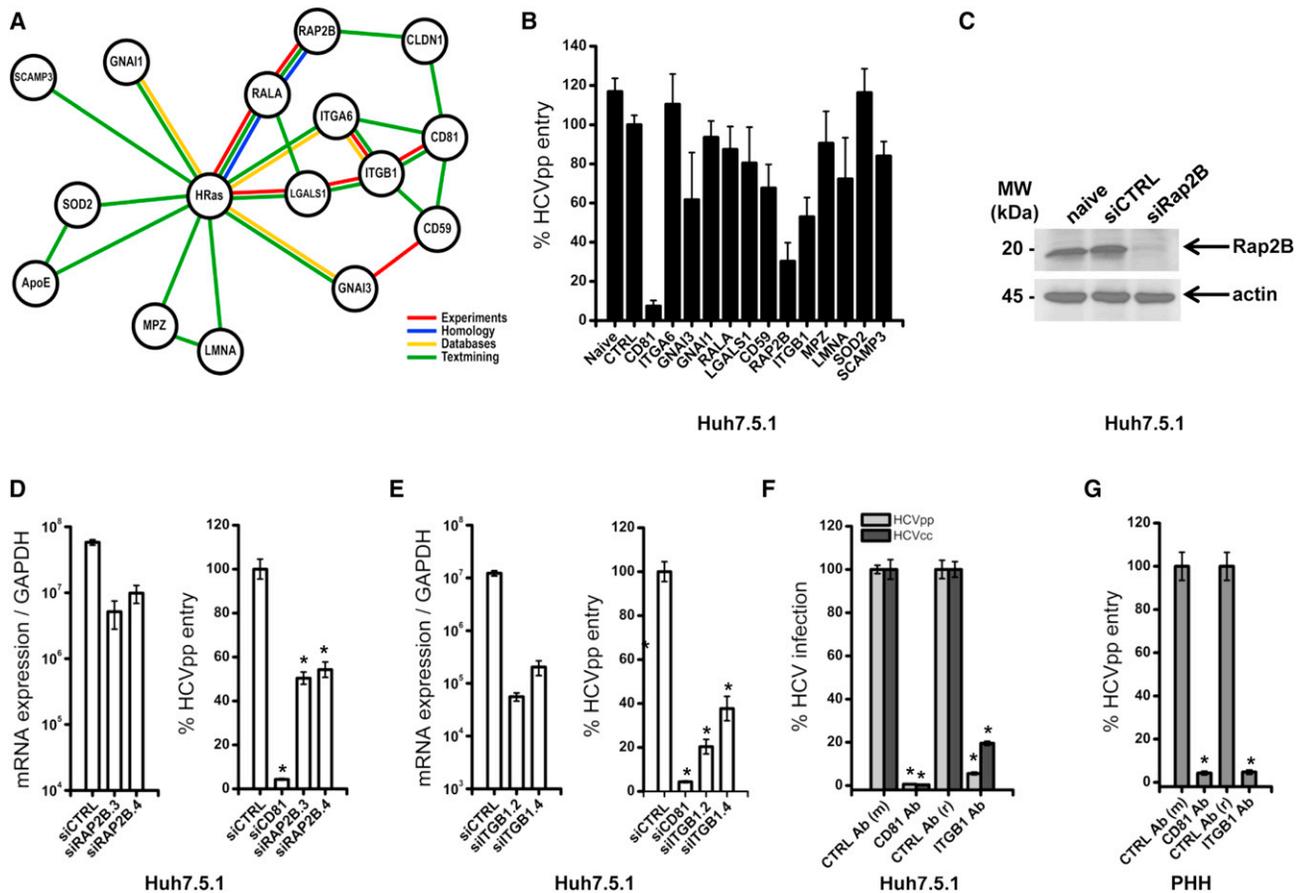
These data lead us to conclude that HCV exploits the EGFR/HRas pathway to compartmentalize host entry factors and receptor trafficking to regulate CD81-dependent pathogen invasion of the liver.

## DISCUSSION

Here, we identified GTPase HRas as key signaling factor in HCV entry. We discovered that HRas associates with CD81-CLDN1, providing a physical link between the EGFR/Shc1/Grb2/HRas signaling pathway and the HCV entry factor complex. Furthermore, we identified the CD81-associated proteins ITGB1 and Rap2B as cofactors for HCV entry. We demonstrate that the identified signaling pathway mediates CD81-CLDN1 coreceptor associations and that HRas signaling regulates CD81 diffusion and confinement in the plasma membrane. Since CD81 lateral diffusion and its association with CLDN1 are essential for HCV entry *in vitro* ([Harris et al., 2012](#); [Harris et al., 2010](#); [Krieger et al., 2010](#)), these findings identify HRas as a trigger of HCV entry.

HCV-CD81 engagement has been suggested to induce MAPK ([Brazzoli et al., 2008](#)) and PI3K/AKT ([Liu et al., 2012](#)) signaling. Moreover, EGFR function is required for HCV entry ([Lupberger et al., 2011](#)) and HCV-CD81 engagement promotes EGFR phosphorylation ([Diao et al., 2012](#)). These studies suggest that HCV may use multiple signaling pathways and mechanisms for entry; however, the underlying molecular mechanisms and the relevance of these pathways for HCV entry into human hepatocytes or liver tissue *in vivo* are unknown. Since transformed hepatoma cells such as Huh7 have deregulated signal transduction pathways, observations might be blurred or confounded by cell-line-specific effects. Here we combined RNAi screening, phosphorylation arrays, and small-molecule inhibitors to study signaling in HCV entry in primary liver cells, concluding a role for EGF priming HRas/MAPK signaling pathway in HCV infection of human hepatocytes and patient-derived liver tissue ([Figures 1 and S1](#)).

Interestingly, although in our hands silencing of PI3K regulatory subunit 1 (PI3KR1) expression reduced HCVpp entry (data not shown) as previously shown ([Liu et al., 2012](#)), we failed to validate its functional relevance for HCV entry using two different small-molecule PI3K inhibitors ([Figures 3G and S2A–S2C](#)).



**Figure 5. Functional Analysis of HRas-CD81-Associated Proteins Identifies Integrin Beta1 and Rap2B as Previously Undiscovered HCV Entry Factors**

(A) Subset of TEM protein association network of the 169 proteins associating with HRas and CD81 identified by STRING analysis (Jensen et al., 2009). Lines connecting proteins show direct (physical) and indirect (functional) associations derived from numerous sources, including experimental repositories (red lines), computational prediction methods (blue lines), databases (yellow lines), and public text collections (green lines).

(B) Functional analysis of protein association network via RNAi. HCVpp entry in Huh7.5.1 cells transfected with pooled siRNA directed against identified members of CD81-associated protein network containing HRas. siCTRL and CD81-specific siRNA served as internal controls. Data are expressed as percentage HCVpp entry relative to siCTRL-transfected cells (means  $\pm$  SD from one representative experiment,  $n = 3$ ).

(C) Analysis of protein expression in lysates of Huh7.5.1 with silenced Rap2B expression by immunoblot with specific antibodies targeting Rap2B or actin (Rap2B pool siRNA).

(D and E) mRNA expression of Rap2B (D) and integrin beta1 (ITGB1) (E) compared to GAPDH or HCVpp entry after silencing of each protein with individual siRNAs. Huh7.5.1 cells were silenced with siRAP2B.3 or siRAP2B.4 and siITGB1.2 or siITGB1.4 for 72 hr prior to mRNA expression measurement or to HCVpp infection of Huh7.5.1 cells. siCTRL and CD81-specific siRNA served as internal controls. Data are expressed as mRNA expression of Rap2B or ITGB1 compared to GAPDH (means  $\pm$  SD from one representative experiment in triplicate,  $n = 3$ ) or percentage HCVpp entry relative to siCTRL-transfected cells (means  $\pm$  SEM from three independent experiments in triplicate,  $n = 9$ ).

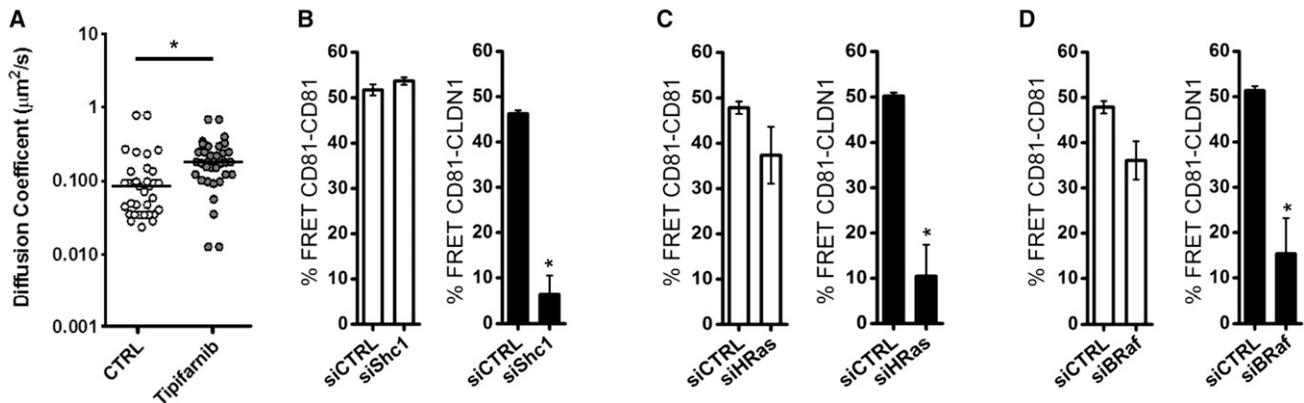
(F and G) Effect of a neutralizing ITGB1-specific antibody on HCV entry and infection. Huh7.5.1 cells (F) or PHHs (G) were treated with 25  $\mu$ g/ml antibodies 1 hr prior and during infection with HCVpp or HCVcc. Irrelevant rabbit and mouse IgGs and a CD81-specific antibody were used as controls. Data are expressed as percentage HCVpp entry or HCVcc infection relative to cells treated with irrelevant IgG (means  $\pm$  SEM from four experiments in triplicate,  $n = 12$ ).

\* $p < 0.0001$ .

Moreover, silencing of Gab1 that binds EGFR and activates PI3K had no significant impact on HCVpp entry (Figure 2), and EGFR signaling through AKT was limited or absent in PHHs or liver tissue in vivo (Figures 1 and S1). Thus, our data suggest that HRas and the upstream MAPK pathway are key signal transducers for EGFR-mediated HCV entry into PHHs and the human liver in vivo and that signal transduction through the PI3K/AKT pathway most likely plays only a minor role.

Our functional analyses suggest that HRas acts as a molecular switch promoting RTK-mediated HCV entry. Inhibition/rescue

experiments highlight that EGFR-mediated HCV entry is dependent on HRas function. Our observation that HRas associates with tetraspanin CD81 supports our biochemical data showing that HRas links RTK signaling to CD81 and promotes CD81-CLDN1 association. This is in line with the recent finding that CD81 internalization via a clathrin- and dynamin-dependent process is independent of the CD81 cytoplasmic domain, suggesting a role for associated partner proteins in regulating CD81 trafficking (Farquhar et al., 2012). Moreover, it has been reported that CD81 engagement activates Rho GTPase family



**Figure 6. HRas Is Required for Lateral Diffusion of CD81 Promoting CD81-CLDN1 Associations**

(A) AcGFP-CD81 diffusion coefficient in Huh7.5.1 cells after 4 hr treatment with DMSO or 10 µM tipifarnib. The median CD81 diffusion coefficient (DMSO, 0.09 µm<sup>2</sup>/s; tipifarnib, 0.18 µm<sup>2</sup>/s) is shown, with each point representing a bleached region of interest and the black line represents the median value.

(B–D) FRET of CD81-CD81 (open bars) and CD81-CLDN1 (black bars) coreceptor associations in Huh7.5.1 cells incubated with siRNA specific for Shc1 (B), HRas (C) or BRaf (D) (means ± SEM from ten independent experiments, n = 10).

\*p < 0.05.

members leading to actin-dependent relocation of HCV E2-CD81 and activation of Raf/MAPK signaling (Brazzoli et al., 2008). Membrane microdomains, such as TEMs or lipid rafts, play a role in a variety of physiological and pathological processes, for instance as signaling platform (Le Naour et al., 2006). TEMs and lipid rafts differ in their solubility in Triton X-100, as well as in their protein composition, and thus are distinct membrane microdomains (Le Naour et al., 2006). GDP-bound inactive HRas is associated to lipid rafts, whereas GTP-bound active HRas is segregated from lipid rafts to bulk plasma membrane microdomains where it activates signal transduction including the Raf/MAPK pathway (Tian et al., 2007). Thus, it is likely that EGFR-induced signals activate HRas function and that GTP-bound activated HRas leads to rearrangement of tetraspanins, resulting in formation of the essential CD81-CLDN1 entry receptor complex.

Our functional analysis of HRas-CD81-associated proteins demonstrates for the first time a functional role of the GTPase Rap2B and ITGB1 as cofactors for HCV entry. Like HRas, Rap GTPases are known regulators of integrin function. Rap increases integrin avidity by promoting integrin clustering (Kinbara et al., 2003) that may have an important impact on the CD81 TEM formation. ITGB1 is a major TEM component. Integrins are heterodimeric transmembrane proteins composed of an alpha and a beta subunit that couple the extracellular matrix to the F-actin cytoskeleton and signal in a bidirectional manner (Wickström and Fässler, 2011). Conformational changes of integrins elicit signaling events that promote cytoskeletal rearrangement and internalization of many viruses (Stewart and Nemerow, 2007). EGFR can be activated in an ITGB1-dependent manner, and ITGB1 controls EGFR signaling (Morello et al., 2011; Moro et al., 1998), suggesting a crosstalk between ITGB1 and EGFR in HCV entry. Collectively, these findings suggest that HRas acts together with Rap2B and ITGB1 to form a functional complex that may regulate host cell entry receptor mobility, as well as plasma membrane and cytoskeleton organization.

Indeed, the HRas inhibitor tipifarnib promotes CD81 lateral diffusion speed, suggesting an inhibitory role for HRas to regulate CD81 diffusion coefficient at the plasma membrane. We previously reported that hepatoma polarization limits CD81 and HCVpp diffusion coefficient (Harris et al., 2012), concluding that CD81 lateral movement plays an essential role in HCV glycoprotein-dependent particle dynamics that are essential for efficient particle entry.

Our results emphasize that TEMs are active and dynamic areas of the membrane and uncover an important role of GTPases as molecular switches to provide a functional link between TEM-associated tetraspanins and the cytoskeleton, allowing efficient coreceptor complex formation and cellular entry of viruses. Indeed, tetraspanins have been associated with the initiation of infection by various pathogens. Moreover, a recent functional siRNA screen has suggested a potential role for CD81 and HRas for influenza virus entry (Karlus et al., 2010), although their exact function in this process was not investigated. Here, we demonstrate that silencing HRas inhibits the entry of pseudoparticles expressing glycoproteins of influenza A and measles virus but not MLV or VSV. Collectively, these findings highlight a functional relevance for HRas and its role in plasma membrane compartmentalization and receptor trafficking for entry of viruses of other families. Furthermore, our results identify a mechanism to regulate CD81-dependent pathogen invasion of the liver that is HRas dependent.

Finally, our results might have therapeutic implications for the treatment of viral infections. Pharmacological interference with BRaf and HRas might provide an approach for fighting a broad range of viral infections including hepatitis C, influenza, and measles. Indeed, host-targeting agents are an emerging strategy to overcome antimicrobial resistance, a major limitation of direct-acting antivirals or antibiotics (Nathan, 2012). The recent development of safe and efficient clinically licensed small-molecule inhibitors of GTPase and BRaf (Downward, 2003; Maurer et al., 2011; Vanneman and Dranoff, 2012) provides a unique opportunity to develop host-targeting antiviral strategies. In

conclusion, our study has important impact not only for the understanding of viral entry and pathogenesis, but also for the development of preventive and therapeutic antiviral strategies.

## EXPERIMENTAL PROCEDURES

### Cell Lines and Primary Human Hepatocytes

The sources and culture conditions for 293T, Huh7, Huh7.5.1, HepG2, and HepG2-CD81 cells have been described (Lupberger et al., 2011; Mee et al., 2009). PHHs were isolated and cultured as described (Krieger et al., 2010; Lupberger et al., 2011). Polarization of HepG2-CD81 was induced as described (Mee et al., 2009).

### Patient-Derived Liver Biopsies

Liver biopsy tissues were analyzed as described (Dill et al., 2012). Liver tissue was lysed and subjected to immunoblot and phosphorylation array analysis (described below). For ex vivo stimulation, liver tissue was incubated for 15 min with 1  $\mu$ g/ml EGF. The protocol was approved by the Ethics Committee of the University Hospital of Basel, Switzerland. Written informed consent was obtained from all patients.

### Analysis of Cell Signaling with Phosphorylation Arrays

Lysates of cells and liver biopsies were subjected to the proteome Profiler Array human phosphokinase array and human phospho-RTK array (R&D Systems) according to the manufacturer's protocol.

### Dynamic Phosphoflow Analyses

Phosphorylated forms of ERK1/ERK2 and AKT were quantified with phospho-specific antibodies in the presence of the phosphatase inhibitor pervanadate, EGF, and kinase inhibitors as described (Fraguay and Nunès, 2009). Cells were fixed, permeabilized, and incubated successively with rabbit antibodies directed against pAKT or pERK1/pERK2, biotinylated anti-rabbit antibodies, and a streptavidin-phycoerythrin solution (Beckman Coulter, Paris, France).

### Infection of Cells with Viral Pseudoparticles and Cell-Culture-Derived HCV

Lentiviral pseudoparticles expressing envelope glycoproteins from HCV (strains HCV-J and P01VL), VSV, MLV, measles, RD114, avian fowl plague influenza A (H7N1), and HCVcc (strain Luc-Jc1) were generated as described (Lupberger et al., 2011). Infection of Huh7.5.1, HepG2-CD81 cells, and PHHs with pseudoparticles and HCVcc were performed as described (Krieger et al., 2010; Lupberger et al., 2011). Unless otherwise stated, pseudoparticle entry and HCVcc infection were assessed by measurement of luciferase activity 72 hr after infection as described (Krieger et al., 2010; Lupberger et al., 2011). HCV cell-cell transmission was assayed as described (Lupberger et al., 2011) and is detailed in the Supplemental Experimental Procedures.

### Functional RNAi HCV Entry Screens

siRNA screens targeting known EGFR binding partners and CD81-associated proteins were applied in Huh7.5.1 cells as described (Lupberger et al., 2011) with ON-TARGETplus smart pools (pools of four individual siRNAs; Dharmacon). For each target, 5.25 pmol siRNA was reverse transfected in 5,000 Huh7.5.1 cells per well of a 96-well microplate with INTERFERin (Polyplus).

### Rescue of EGFR Inhibition with a HRas Transcomplementation Assay

Huh7.5.1 cells ( $0.66 \times 10^4$ ) or PHHs were seeded as described (Lupberger et al., 2011) 1 day prior to transduction with lentiviruses expressing the transdominant active HRas V12 mutant or control (Beauséjour et al., 2003). Seventy-two hours later, cells were infected with HCVpp in the presence of 10  $\mu$ M erlotinib or DMSO control.

### Proteomic Analyses of Tetraspanin Complexes and Microdomains

SILAC was performed as described (Ong et al., 2002). HepG2-CD81 cells and control HepG2 were incubated with either light or heavy isotope labeled amino acids and then lysed with brij97. The two lysates were pooled, and CD81-associated proteins were coimmunoprecipitated as described (André

et al., 2006). The proteins were separated by SDS-PAGE and identified by liquid chromatography-mass spectrometry. A peak volume ratio heavy/light >2 was defined as threshold for potential CD81-associated proteins. More details are given in the Supplemental Experimental Procedures.

### Imaging Studies

FRAP was performed as described (Harris et al., 2012). Huh7.5.1 cells were transduced with GFP-labeled CD81 (AcGFP-CD81), and CD81 motility was assessed at the membrane of live cells with TIRF microscopy after photobleaching. FRET analyses of homotypic and heterotypic interactions of CD81 and CLDN1 were analyzed in Huh7.5.1 cells as described (Harris et al., 2010). The data from ten cells were normalized, and the localized expression was calculated. Confocal microscopy and staining was performed as described (Chambrion and Le Naour, 2010). Colocalization was calculated according to Pearson and Manders (Bolte and Cordelières, 2006; Manders et al., 1992). More details are given in the Supplemental Experimental Procedures.

### Statistical Analysis

All experiments were performed at least three times in triplicate in an independent manner, and results were analyzed with the nonparametric Mann-Whitney test if not indicated otherwise. An F test was performed for analysis of variance (one-way ANOVA) of colocalization studies to compare means of two groups ( $n = 3$ ) of Pearson's correlation coefficient (Rr) or ICQ. Significant p values are indicated by an asterisk in the individual figure legends.

Additional information on experimental procedures is provided in the Supplemental Experimental Procedures.

## SUPPLEMENTAL INFORMATION

Supplemental Information includes four figures, one table, and Supplemental Experimental Procedures and can be found with this article online at <http://dx.doi.org/10.1016/j.chom.2013.02.006>.

## ACKNOWLEDGMENTS

This work was supported by the European Union (ERC-2008-AdG-233130-HEPCENT and INTERREG-IV-Rhin Supérieur-FEDER A9 and A30), Inserm, Laboratoire d'Excellence HEPYS (ANR-10-LAB-28), ANRS (2010/304, 2011/132, 2012/239), and DGOS; J.A.M. acknowledges funding from the Wellcome Trust. The authors thank P. Pessaux and P. Bachelier (Strasbourg University Hospitals) for liver resections; R. Bartenschlager (University of Heidelberg, Heidelberg, Germany) for HCV strain Luc-Jc1; F.V. Chisari (The Scripps Research Institute, La Jolla, CA), C.M. Rice (Rockefeller University, New York, NY), and A. Patel (University of Glasgow, Glasgow, UK) for Huh7.5.1 and Huh7.5-GFP cells; E. Campeau and C. Beauséjour for HRas V12 plasmid; D. Trono for pWPI plasmid; and C. Schuster (Inserm U1110, Strasbourg, France) for HepG2-CD81 cells. We thank A. Weiss, C. Bach, and M. Parnot for excellent technical work and Y. Couté for his helpful contributions to the SILAC analysis. L.Z. and J.L. as well as N.S.-A. and C.T. contributed equally. L.Z., J.L., F.L.N., D.S., I.H., H.J.H., A.B., M.B.Z., J.A.M., L.B., F.H.T.D., M.H.H., and T.F.B. designed experiments. L.Z., J.L., C.T., H.J.H., A.B., N.S.-A., R.G.T., F.X., M.T., S.C.D., J.F., and F.H.T.D. performed experiments. F.L.C. provided essential reagents. L.Z., J.L., F.L.N., N.S.-A., D.S., I.H., H.J.H., J.A.M., L.B., F.H.T.D., M.H.H., M.B.Z., and T.F.B. analyzed data. L.Z., J.L., F.L.N., M.B.Z., J.A.M., and T.F.B. wrote the manuscript.

Received: August 2, 2012

Revised: January 3, 2013

Accepted: February 11, 2013

Published: March 13, 2013

## REFERENCES

André, M., Le Caer, J.P., Greco, C., Planchon, S., El Nemer, W., Boucheix, C., Rubinstein, E., Chamot-Rooke, J., and Le Naour, F. (2006). Proteomic analysis of the tetraspanin web using LC-ESI-MS/MS and MALDI-FTICR-MS. *Proteomics* 6, 1437–1449.

- Barth, H., Schnober, E.K., Zhang, F., Linhardt, R.J., Depla, E., Boson, B., Cosset, F.L., Patel, A.H., Blum, H.E., and Baumert, T.F. (2006). Viral and cellular determinants of the hepatitis C virus envelope-heparan sulfate interaction. *J. Virol.* **80**, 10579–10590.
- Beauséjour, C.M., Krtolica, A., Galimi, F., Narita, M., Lowe, S.W., Yaswen, P., and Campisi, J. (2003). Reversal of human cellular senescence: roles of the p53 and p16 pathways. *EMBO J.* **22**, 4212–4222.
- Berdichevski, F., and Odintsova, E. (2007). Tetraspanins as regulators of protein trafficking. *Traffic* **8**, 89–96.
- Boguski, M.S., and McCormick, F. (1993). Proteins regulating Ras and its relatives. *Nature* **366**, 643–654.
- Bolte, S., and Cordelières, F.P. (2006). A guided tour into subcellular colocalization analysis in light microscopy. *J. Microsc.* **224**, 213–232.
- Brazzoli, M., Bianchi, A., Filippini, S., Weiner, A., Zhu, Q., Pizza, M., and Crotta, S. (2008). CD81 is a central regulator of cellular events required for hepatitis C virus infection of human hepatocytes. *J. Virol.* **82**, 8316–8329.
- Brown, M., Adyshev, D., Bindokas, V., Moitra, J., Garcia, J.G., and Dudek, S.M. (2010). Quantitative distribution and colocalization of non-muscle myosin light chain kinase isoforms and cortactin in human lung endothelium. *Microvasc. Res.* **80**, 75–88.
- Chakraborty, S., Veettil, M.V., Bottero, V., and Chandran, B. (2012). Kaposi's sarcoma-associated herpesvirus interacts with EphrinA2 receptor to amplify signaling essential for productive infection. *Proc. Natl. Acad. Sci. USA* **109**, E1163–E1172.
- Chambrion, C., and Le Naour, F. (2010). The tetraspanins CD9 and CD81 regulate CD9P1-induced effects on cell migration. *PLoS ONE* **5**, e11219.
- Diao, J., Pantua, H., Ngu, H., Komuves, L., Diehl, L., Schaefer, G., and Kapadia, S.B. (2012). Hepatitis C virus induces epidermal growth factor receptor activation via CD81 binding for viral internalization and entry. *J. Virol.* **86**, 10935–10949.
- Dill, M.T., Makowska, Z., Duong, F.H., Merkofer, F., Filipowicz, M., Baumert, T.F., Tornillo, L., Terracciano, L., and Heim, M.H. (2012). Interferon- $\gamma$ -stimulated genes, but not USP18, are expressed in livers of patients with acute hepatitis C. *Gastroenterology* **143**, 777–786, e1–e6.
- Downward, J. (2003). Targeting RAS signalling pathways in cancer therapy. *Nat. Rev. Cancer* **3**, 11–22.
- El-Serag, H.B. (2012). Epidemiology of viral hepatitis and hepatocellular carcinoma. *Gastroenterology* **142**, 1264–1273, e1.
- Evans, M.J., von Hahn, T., Tschernie, D.M., Syder, A.J., Panis, M., Wölk, B., Hatzioannou, T., McKeating, J.A., Bieniasz, P.D., and Rice, C.M. (2007). Claudin-1 is a hepatitis C virus co-receptor required for a late step in entry. *Nature* **446**, 801–805.
- Farquhar, M.J., Hu, K., Harris, H.J., Davis, C., Brimacombe, C.L., Fletcher, S.J., Baumert, T.F., Rappoport, J.Z., Balfe, P., and McKeating, J.A. (2012). Hepatitis C virus induces CD81 and claudin-1 endocytosis. *J. Virol.* **86**, 4305–4316.
- Firaguy, G., and Nunès, J.A. (2009). Analysis of signaling events by dynamic phosphoflow cytometry. *Sci. Signal.* **2**, p13.
- Harris, H.J., Davis, C., Mullins, J.G., Hu, K., Goodall, M., Farquhar, M.J., Mee, C.J., McCaffrey, K., Young, S., Drummer, H., et al. (2010). Claudin association with CD81 defines hepatitis C virus entry. *J. Biol. Chem.* **285**, 21092–21102.
- Harris, H.J., Clerte, C., Farquhar, M.J., Goodall, M., Hu, K., Rassam, P., Dosset, P., Wilson, G.K., Balfe, P., Ijzendoorn, S.C., et al. (2012). Hepatoma polarization limits CD81 and hepatitis C virus dynamics. *Cell. Microbiol.* **15**, 430–445.
- Jensen, L.J., Kuhn, M., Stark, M., Chaffron, S., Creevey, C., Muller, J., Doerks, T., Julien, P., Roth, A., Simonovic, M., et al. (2009). STRING 8—a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res.* **37**(Database issue), D412–D416.
- Karlas, A., Machuy, N., Shin, Y., Pleissner, K.P., Artarini, A., Heuer, D., Becker, D., Khalil, H., Ogilvie, L.A., Hess, S., et al. (2010). Genome-wide RNAi screen identifies human host factors crucial for influenza virus replication. *Nature* **463**, 818–822.
- Kinbara, K., Goldfinger, L.E., Hansen, M., Chou, F.L., and Ginsberg, M.H. (2003). Ras GTPases: integrins' friends or foes? *Nat. Rev. Mol. Cell Biol.* **4**, 767–776.
- Kolch, W. (2005). Coordinating ERK/MAPK signalling through scaffolds and inhibitors. *Nat. Rev. Mol. Cell Biol.* **6**, 827–837.
- König, R., Stertz, S., Zhou, Y., Inoue, A., Hoffmann, H.H., Bhattacharyya, S., Alamares, J.G., Tschernie, D.M., Ortigoza, M.B., Liang, Y., et al. (2010). Human host factors required for influenza virus replication. *Nature* **463**, 813–817.
- Kremontsov, D.N., Rassam, P., Margeat, E., Roy, N.H., Schneider-Schaulies, J., Milhiet, P.E., and Thali, M. (2010). HIV-1 assembly differentially alters dynamics and partitioning of tetraspanins and raft components. *Traffic* **11**, 1401–1414.
- Krieger, S.E., Zeisel, M.B., Davis, C., Thumann, C., Harris, H.J., Schnober, E.K., Mee, C., Soulier, E., Royer, C., Lambotin, M., et al. (2010). Inhibition of hepatitis C virus infection by anti-claudin-1 antibodies is mediated by neutralization of E2-CD81-claudin-1 associations. *Hepatology* **51**, 1144–1157.
- Le Naour, F., André, M., Boucheix, C., and Rubinstein, E. (2006). Membrane microdomains and proteomics: lessons from tetraspanin microdomains and comparison with lipid rafts. *Proteomics* **6**, 6447–6454.
- Liu, S., Yang, W., Shen, L., Turner, J.R., Coyne, C.B., and Wang, T. (2009). Tight junction proteins claudin-1 and occludin control hepatitis C virus entry and are downregulated during infection to prevent superinfection. *J. Virol.* **83**, 2011–2014.
- Liu, Z., Tian, Y., Machida, K., Lai, M.M., Luo, G., Fong, S.K., and Ou, J.H. (2012). Transient activation of the PI3K-AKT pathway by hepatitis C virus to enhance viral entry. *J. Biol. Chem.* **287**, 41922–41930.
- Lupberger, J., Zeisel, M.B., Xiao, F., Thumann, C., Fofana, I., Zona, L., Davis, C., Mee, C.J., Turek, M., Gorke, S., et al. (2011). EGFR and EphA2 are host factors for hepatitis C virus entry and possible targets for antiviral therapy. *Nat. Med.* **17**, 589–595.
- Manders, E.M., Stap, J., Brakenhoff, G.J., van Driel, R., and Aten, J.A. (1992). Dynamics of three-dimensional replication patterns during the S-phase, analysed by double labelling of DNA and confocal microscopy. *J. Cell Sci.* **103**, 857–862.
- Maurer, G., Tarkowski, B., and Baccarini, M. (2011). Raf kinases in cancer: roles and therapeutic opportunities. *Oncogene* **30**, 3477–3488.
- Mee, C.J., Harris, H.J., Farquhar, M.J., Wilson, G., Reynolds, G., Davis, C., van Ijzendoorn, S.C., Balfe, P., and McKeating, J.A. (2009). Polarization restricts hepatitis C virus entry into HepG2 hepatoma cells. *J. Virol.* **83**, 6211–6221.
- Mercer, J., and Helenius, A. (2008). Vaccinia virus uses macropinocytosis and apoptotic mimicry to enter host cells. *Science* **320**, 531–535.
- Morandell, S., Stasyk, T., Skvortsov, S., Ascher, S., and Huber, L.A. (2008). Quantitative proteomics and phosphoproteomics reveal novel insights into complexity and dynamics of the EGFR signaling network. *Proteomics* **8**, 4383–4401.
- Morello, V., Cabodi, S., Sigismund, S., Camacho-Leal, M.P., Repetto, D., Volante, M., Papotti, M., Turco, E., and Defilippi, P. (2011).  $\beta$ 1 integrin controls EGFR signaling and tumorigenic properties of lung cancer cells. *Oncogene* **30**, 4087–4096.
- Moro, L., Venturino, M., Bozzo, C., Silengo, L., Altruda, F., Beguinot, L., Tarone, G., and Defilippi, P. (1998). Integrins induce activation of EGF receptor: role in MAP kinase induction and adhesion-dependent cell survival. *EMBO J.* **17**, 6622–6632.
- Murray, C.L., and Rice, C.M. (2011). Turning hepatitis C into a real virus. *Annu. Rev. Microbiol.* **65**, 307–327.
- Nathan, C. (2012). Fresh approaches to anti-infective therapies. *Sci. Transl. Med.* **4**, sr2.
- Ong, S.E., Blagoev, B., Kratchmarova, I., Kristensen, D.B., Steen, H., Pandey, A., and Mann, M. (2002). Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics* **1**, 376–386.

- Pelkmans, L., Fava, E., Grabner, H., Hannus, M., Habermann, B., Krausz, E., and Zerial, M. (2005). Genome-wide analysis of human kinases in clathrin- and caveolae/raft-mediated endocytosis. *Nature* *436*, 78–86.
- Pileri, P., Uematsu, Y., Campagnoli, S., Galli, G., Falugi, F., Petracca, R., Weiner, A.J., Houghton, M., Rosa, D., Grandi, G., and Abrignani, S. (1998). Binding of hepatitis C virus to CD81. *Science* *282*, 938–941.
- Ploss, A., Evans, M.J., Gaysinskaya, V.A., Panis, M., You, H., de Jong, Y.P., and Rice, C.M. (2009). Human occludin is a hepatitis C virus entry factor required for infection of mouse cells. *Nature* *457*, 882–886.
- Sainz, B., Jr., Barretto, N., Martin, D.N., Hiraga, N., Imamura, M., Hussain, S., Marsh, K.A., Yu, X., Chayama, K., Alrefai, W.A., and Uprichard, S.L. (2012). Identification of the Niemann-Pick C1-like 1 cholesterol absorption receptor as a new hepatitis C virus entry factor. *Nat. Med.* *18*, 281–285.
- Scarselli, E., Ansuini, H., Cerino, R., Roccasecca, R.M., Acali, S., Filocamo, G., Traboni, C., Nicosia, A., Cortese, R., and Vitelli, A. (2002). The human scavenger receptor class B type I is a novel candidate receptor for the hepatitis C virus. *EMBO J.* *21*, 5017–5025.
- Silvie, O., Rubinstein, E., Franetich, J.F., Prenant, M., Belnoue, E., Rénia, L., Hannoun, L., Eling, W., Levy, S., Boucheix, C., and Mazier, D. (2003). Hepatocyte CD81 is required for *Plasmodium falciparum* and *Plasmodium yoelii* sporozoite infectivity. *Nat. Med.* *9*, 93–96.
- Stewart, P.L., and Nemerow, G.R. (2007). Cell integrins: commonly used receptors for diverse viral pathogens. *Trends Microbiol.* *15*, 500–507.
- Tian, T., Harding, A., Inder, K., Plowman, S., Parton, R.G., and Hancock, J.F. (2007). Plasma membrane nanoswitches generate high-fidelity Ras signal transduction. *Nat. Cell Biol.* *9*, 905–914.
- Vanneman, M., and Dranoff, G. (2012). Combining immunotherapy and targeted therapies in cancer treatment. *Nat. Rev. Cancer* *12*, 237–251.
- Wickström, S.A., and Fässler, R. (2011). Regulation of membrane traffic by integrin signaling. *Trends Cell Biol.* *21*, 266–273.
- Zeisel, M.B., Fofana, I., Fafi-Kremer, S., and Baumert, T.F. (2011). Hepatitis C virus entry into hepatocytes: molecular mechanisms and targets for antiviral therapies. *J. Hepatol.* *54*, 566–576.

# Circular RNAs are a large class of animal RNAs with regulatory potency

Sebastian Memczak<sup>1\*</sup>, Marvin Jens<sup>1\*</sup>, Antigoni Elefantioti<sup>1\*</sup>, Francesca Torti<sup>1\*</sup>, Janna Krueger<sup>2</sup>, Agnieszka Rybak<sup>1</sup>, Luisa Maier<sup>1</sup>, Sebastian D. Mackowiak<sup>1</sup>, Lea H. Gregersen<sup>3</sup>, Mathias Munschauer<sup>3</sup>, Alexander Loewer<sup>4</sup>, Ulrike Ziebold<sup>1</sup>, Markus Landthaler<sup>3</sup>, Christine Kocks<sup>1</sup>, Ferdinand le Noble<sup>5</sup> & Nikolaus Rajewsky<sup>1</sup>

Circular RNAs (circRNAs) in animals are an enigmatic class of RNA with unknown function. To explore circRNAs systematically, we sequenced and computationally analysed human, mouse and nematode RNA. We detected thousands of well-expressed, stable circRNAs, often showing tissue/developmental-stage-specific expression. Sequence analysis indicated important regulatory functions for circRNAs. We found that a human circRNA, antisense to the cerebellar degeneration-related protein 1 transcript (CDR1as), is densely bound by microRNA (miRNA) effector complexes and harbours 63 conserved binding sites for the ancient miRNA miR-7. Further analyses indicated that CDR1as functions to bind miR-7 in neuronal tissues. Human CDR1as expression in zebrafish impaired midbrain development, similar to knocking down miR-7, suggesting that CDR1as is a miRNA antagonist with a miRNA-binding capacity ten times higher than any other known transcript. Together, our data provide evidence that circRNAs form a large class of post-transcriptional regulators. Numerous circRNAs form by head-to-tail splicing of exons, suggesting previously unrecognized regulatory potential of coding sequences.

Mature messenger RNAs are linear molecules with 5' and 3' termini that reflect start and stop of the RNA polymerase on the DNA template. In cells, different RNA molecules are sometimes joined together by splicing reactions (trans-splicing), but covalent linkage of the ends of a single RNA molecule to form a circular RNA (circRNA) is usually considered a rare event. circRNAs were discovered in plants and shown to encode subviral agents<sup>1</sup>. In unicellular organisms, circRNAs mostly stem from self-splicing introns of pre-ribosomal RNA<sup>2</sup>, but can also arise from protein-coding genes in archaea<sup>3</sup>. In the few unambiguously validated circRNAs in animals, the spliceosome seems to link the 5' and downstream 3' ends of exons within the same transcript<sup>4–10</sup>. Perhaps the best known circRNA is antisense to the mRNA transcribed from the *SRY* (sex-determining region Y) locus and is highly expressed in testes<sup>6</sup>. Evidence from computational analyses of expression data in Archaea and Mammalia suggests that circRNAs are more prevalent than previously thought<sup>3,10</sup>; however, it is unknown whether animal circRNAs have any biological function.

In comparison to circRNAs, miRNAs are extremely well studied. miRNAs are ~21-nucleotide-long non-coding RNAs that guide the effector protein Argonaute (AGO) to mRNAs of coding genes to repress their protein production<sup>11–14</sup>. In humans, miRNAs directly regulate expression of most mRNAs<sup>15–18</sup> in a diverse range of biological functions. However, surprisingly little is known about how and if mRNAs can escape regulation by a miRNA. A recently discovered mechanism for miRNA removal in a sequence-specific manner is based on target sites acting as decoys or miRNA sponges<sup>19,20</sup>. RNA with miRNA binding sites should, if expressed highly enough, sequester away the miRNA from its target sites. However, all reported mammalian miRNA sponges have only one or two binding sites for the same miRNA and are not highly expressed, limiting their potency<sup>21–24</sup>.

To identify circRNAs across animal cells systematically, we screened RNA-seq data for circRNAs. Compared to previous approaches<sup>10</sup> our

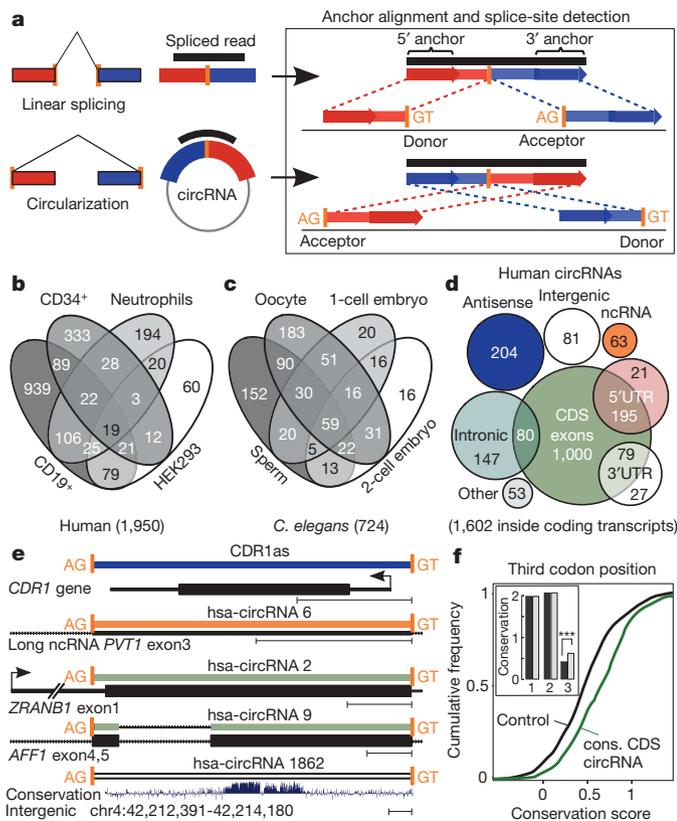
computational pipeline can find circRNAs in any genomic region, takes advantage of long (~100 nucleotides) reads, and predicts the acceptor and donor splice sites used to link the ends of the RNAs. We do not rely on paired-end sequencing data or known splice sites. Using published<sup>10,25,26</sup> and our own sequencing data, our method reported thousands of circRNAs in human and mouse tissues as well as in different developmental stages of *Caenorhabditis elegans*. Numerous circRNAs appear to be specifically expressed across tissues or developmental stages. We validated these data and showed that most tested circRNAs are well expressed, stable and circularized using the predicted splice sites. circRNA sequences were significantly enriched in conserved nucleotides, indicating that circRNAs compete with other RNAs for binding by RNA binding proteins (RBPs) or miRNAs. We combined biochemical, functional and computational analyses to show that indeed a known human circRNA, *CDR1* antisense (CDR1as)<sup>9</sup>, can function as a negative regulator of miR-7, a miRNA with perfect sequence conservation from annelids to human. Together, our data provide evidence that circRNAs form an important class of post-transcriptional regulators.

## circRNAs have complex expression patterns

To comprehensively identify stably expressed circRNAs in animals we screened RNA sequencing reads for splice junctions formed by an acceptor splice site at the 5' end of an exon and a donor site at a downstream 3' end (head-to-tail) (Fig. 1a). As standard RNA expression profiling enriches for polyadenylated RNAs, we used data generated after ribosomal RNA depletion (ribominus) and random priming. Such data were used before to detect scrambled exons in mammals<sup>10</sup> (see Methods for comparison). However, this approach was not specifically designed to detect circRNAs and (1) only used existing exon-intron annotations, thus missing RNAs transcribed from introns or unannotated transcripts; (2) did not explicitly identify

<sup>1</sup>Systems Biology of Gene Regulatory Elements, Max-Delbrück-Center for Molecular Medicine, Robert-Rössle-Strasse 10, 13125 Berlin, Germany. <sup>2</sup>Angiogenesis and Cardiovascular Pathology, Max-Delbrück-Center for Molecular Medicine, Robert-Rössle-Strasse 10, 13125 Berlin, Germany. <sup>3</sup>RNA Biology and Post-Transcriptional Regulation, Max-Delbrück-Center for Molecular Medicine, Robert-Rössle-Strasse 10, 13125 Berlin, Germany. <sup>4</sup>Signaling Dynamics in Single Cells, Max-Delbrück-Center for Molecular Medicine, Robert-Rössle-Strasse 10, 13125 Berlin, Germany.

\*These authors contributed equally to this work.



**Figure 1 | Detection, classification and evolutionary conservation of circRNAs.** **a**, The termini of junction-spanning reads (anchors) align sequentially to the genome for linear (top) but in reversed orientation for head-to-tail spliced reads (bottom). Spliced reads must distribute completely to anchors, flanked by AG/GU (Methods). **b**, **c**, circRNAs in human cell types (**b**) and nematode stages (**c**). **d**, Genomic origin of human circRNAs. A total of 96% of circRNAs overlap known transcripts. **e**, Examples of human circRNAs. The *AFF1* intron is spliced out (Supplementary Fig. 2e). Sequence conservation: placental mammals phyloP score (Methods), scale bar, 200 nucleotides. **f**, A total of 223 human coding sequence circRNAs with mouse orthologues (green) and controls (black) with matched conservation level (inset: mean conservation for each codon position (grey), controls (black); *x* axis, codon positions; *y* axis, placental mammals phyloP score; see also Methods and Supplementary Fig. 1j, k). Third codon positions are significantly more conserved ( $P < 4 \times 10^{-10}$ , Mann–Whitney *U*-test,  $n = 223$ ).

the splice sites used for circularization; and (3) assumed that each pair of mates in paired-end sequencing derives from the same RNA molecule. To search in a more unbiased way for circRNAs, we designed an algorithm (Methods) that identifies linear and circular splicing events in ribominus data. First, we filtered out reads that aligned contiguously to the genome, retaining the spliced reads. Next, we mapped the terminal parts of each candidate read independently to the genome to find unique anchor positions. Finally, we demanded that (1) anchor alignments can be extended such that the original read sequence aligns completely, and (2) the inferred breakpoint is flanked by GU/AG splice signals. Non-unique mappings and ambiguous breakpoints were discarded. We detected circularization splicing from the reversed (head-to-tail) orientation of the anchor alignments (Fig. 1a). Our method also recovered tens of thousands of known linear splicing events (Methods and Supplementary Fig. 1a, b). We estimated sensitivity (>75%) and false-discovery rate (FDR <0.2%) using simulated reads and various permutations of real sequencing data (Methods and Supplementary Fig. 1c). However, the efficiency of ribominus protocols to extract and sequence circRNAs is limited, reducing overall sensitivity.

We generated ribominus data for HEK293 cells and, combined with human leukocyte data<sup>10</sup>, detected 1,950 circRNAs with support

from at least two independent junction-spanning reads (Fig. 1b). The expression of genes predicted to give rise to circRNAs was only slightly shifted towards higher expression values (Supplementary Fig. 1d), indicating that circRNAs are not just rare mistakes of the spliceosome. We also identified 1,903 circRNAs in mouse (brains, fetal head, differentiation-induced embryonic stem cells; Supplementary Fig. 1e)<sup>25,26</sup>; 81 of these mapped to human circRNAs (Supplementary Fig. 1f). To explore whether circRNAs exist in other animal clades, we used sequencing data that we produced from various *C. elegans* developmental stages (Stoeckius, M. *et al.*, manuscript in preparation) (Methods) and detected 724 circRNAs, with at least two independent reads (Fig. 1c).

Numerous circRNAs seem to be specifically expressed in a cell type or developmental stage (Fig. 1b, c and Supplementary Fig. 1e). For example, hsa-circRNA 2149 is supported by 13 unique, head-to-tail spanning reads in CD19<sup>+</sup> leukocytes but is not detected in CD34<sup>+</sup> leukocytes (which were sequenced at comparable depth; Supplementary Table 1), neutrophils or HEK293 cells. Analogously, a number of nematode circRNAs seem to be expressed in oocytes but absent in 1- or 2-cell embryos.

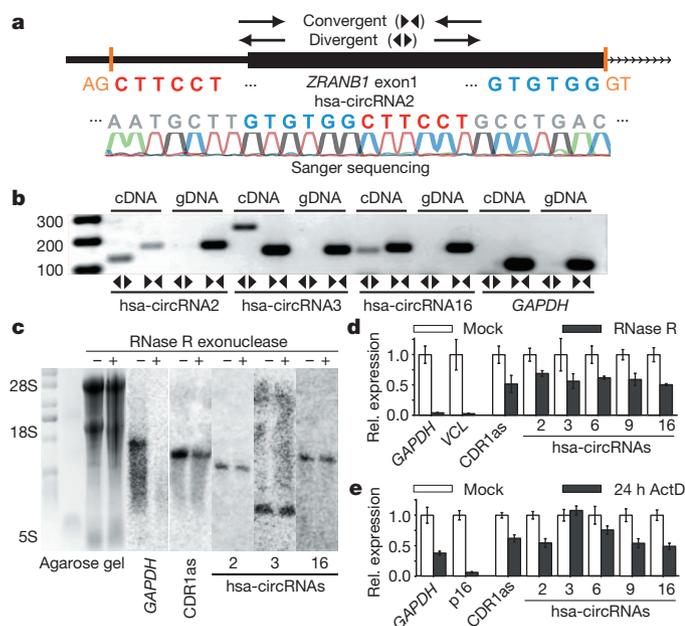
We annotated human circRNAs using the RefSeq database and a catalogue of non-coding RNAs<sup>27–29</sup>. 85% of human circRNAs align sense to known genes. Their splice sites typically span one to five exons (Supplementary Fig. 1g) and overlap coding exons (84%), but only in 65% of these cases are both splice sites that participate in the circularization known splice sites (Supplementary Table 2), demonstrating the advantage of our strategy. 10% of all circRNAs align antisense to known transcripts, smaller fractions align to UTRs, introns, unannotated regions of the genome (Fig. 1d). Examples of human circRNAs are shown in Fig. 1e.

We analysed sequence conservation within circRNAs. As genomic sequence is subject to different degrees of evolutionary selection, depending on function, we studied three subtypes of circRNAs. Intergenic and a few intronic circRNAs display a mild but significant enrichment of conserved nucleotides (Supplementary Fig. 1h, i). To analyse circRNAs composed of coding sequence and thus high overall conservation, we selected 223 human circRNAs with circular orthologues in mouse (Methods) and entirely composed of coding sequence. Control (linear) exons were randomly selected to match the level of conservation observed in first and second codon positions (Methods, Fig. 1f inset and Supplementary Fig. 1k for conservation of the remaining coding sequence (CDS)). circRNAs with conserved circularization were significantly more conserved in the third codon position than controls, indicating evolutionary constraints at the nucleotide level, in addition to selection at the protein level (Fig. 1f and Supplementary Fig. 1j, k). In summary, we have confidently identified a large number of circRNAs with complex expression patterns, which derive often but not always from coding exons. Sequence conservation suggests that at least a subset contains functional sequence elements.

### Characterization of 50 predicted circRNAs

We experimentally tested our circRNA predictions in HEK293 cells. Head-to-tail splicing was assayed by quantitative polymerase chain reaction (qPCR) after reverse transcription, with divergent primers and Sanger sequencing (Fig. 2a, b). Predicted head-to-tail junctions of 19 out of 23 randomly chosen circRNAs (83%) could be validated, demonstrating high accuracy of our predictions (Table 1). In contrast, 5 out of 7 (71%) candidates exclusively predicted in leukocytes could not be detected in HEK293 cells, validating cell-type-specific expression.

Head-to-tail splicing could be produced by trans-splicing or genomic rearrangements. To rule out these possibilities as well as potential PCR artefacts, we successfully validated the insensitivity of human circRNA candidates to digestion with RNase R—an exonuclease that degrades linear RNA molecules<sup>30</sup>—by northern blotting with probes which span the head-to-tail junctions (Fig. 2c). We quantified RNase R resistance for 21 candidates with confirmed head-to-tail splicing by



**Figure 2 | CircRNAs are stable transcripts with robust expression.** **a**, Human (hsa) *ZRANB1* circRNA exemplifies the validation strategy. Convergent (divergent) primers detect total (circular) RNAs. Sanger sequencing confirms head-to-tail splicing. **b**, Divergent primers amplify circRNAs in cDNA but not genomic DNA (gDNA). *GAPDH*, linear control, size marker in base pairs. **c**, Northern blots of mock (–) and RNase R (+) treated HEK293 total RNA with head-to-tail specific probes for circRNAs. *GAPDH*, linear control. **d**, **e**, circRNAs are at least 10-fold more RNase R resistant than *GAPDH* mRNA (**d**) and stable after 24 h transcription block (**e**) (qPCR; error bars indicate standard deviation).

qPCR. All of these were at least 10-fold more resistant than *GAPDH* (Fig. 2d and Supplementary Fig. 2a). We reasoned that circRNAs should generally turn over more slowly than mRNAs. Indeed, we found that 24 h after blocking transcription circRNAs were highly stable, exceeding the stability of the housekeeping gene *GAPDH*<sup>31</sup> (Fig. 2e and Supplementary Fig. 2b). We also validated 3 out of 3 tested mouse circRNAs with human orthologues in mouse brains (Supplementary Fig. 2c). In *C. elegans* 15 out of 20 (75%) of the predictions from gametes and early embryos were validated in a mixed stage sample (Supplementary Fig. 2d and Supplementary Table 3).

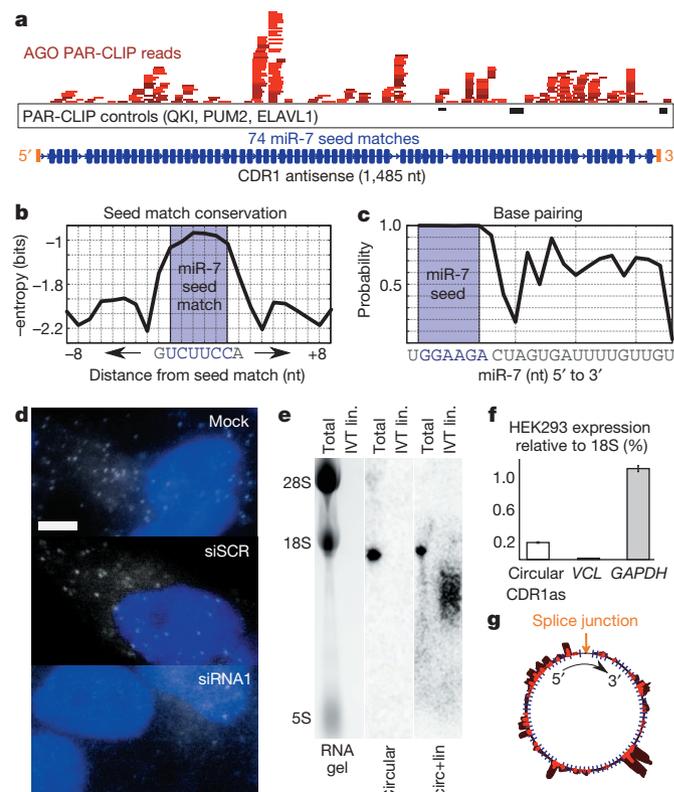
### circRNA CDR1as is densely bound by AGO

Stable transcripts with many miRNA-binding sites could function as miRNA sponges. We intersected our catalogue of circRNAs with transcript annotations, assuming that introns would not occur in mature circRNAs (as observed for 3 out of 3 tested circRNAs, Supplementary Fig. 2e). We screened for occurrences of conserved miRNA family seed matches (Methods). When counting repetitions of conserved matches to the same miRNA family, circRNAs were

significantly enriched compared to coding sequences ( $P < 2.96 \times 10^{-22}$ , Mann-Whitney *U*-test,  $n = 3,873$ ) or 3' UTR sequences ( $P < 2.76 \times 10^{-21}$ , Mann-Whitney *U*-test,  $n = 3,182$ ) (Supplementary Fig. 3a, b).

As an extreme case, we discovered that the known human circRNA CDR1as (ref. 9) harboured dozens of conserved miR-7 seed matches. To test whether CDR1as is bound by miRNAs, we analysed biochemical, transcriptome-wide binding-site data for the miRNA effector AGO proteins. We performed four independent PAR-CLIP (photoactivatable-ribonucleoside-enhanced crosslinking and immunoprecipitation) experiments for human AGO (Methods) and analysed them together with published, lower-depth data<sup>32</sup>. PAR-CLIP<sup>32–34</sup> is based on ultraviolet crosslinking of RNA to protein and subsequent sequencing of RNA bound to a RBP of interest. The ~1.5-kilobase (kb) *CDR1as* locus stood out in density and number of AGO PAR-CLIP reads (Fig. 3a), whereas nine combined PAR-CLIP libraries for other RBPs gave virtually no signal. Of note, there is no PAR-CLIP read mapping to the sense coding transcript of the *CDR1* gene, which was originally identified as a target of autoantibodies from patients with paraneoplastic cerebellar degeneration<sup>35</sup>.

Sequence analysis across 32 vertebrate species revealed that miR-7 is the only animal miRNA with conserved seed matches that can explain the AGO binding along the CDR1as transcript (Methods). Human CDR1as harbours 74 miR-7 seed matches of which 63 are



**Figure 3 | The circRNA CDR1as is bound by the miRNA effector protein AGO, and is cytoplasmic.** **a**, CDR1as is densely bound by AGO (red) but not by unrelated proteins (black). Blue boxes indicate miR-7 seed matches. nt, nucleotides. **b**, **c**, miR-7 sites display reduced nucleotide variability across 32 vertebrate genomes (**b**) and high base-pairing probability within seed matches (**c**). **d**, CDR1as RNA is cytoplasmic and disperse (white spots; single-molecule RNA FISH; maximum intensity merges of Z-stacks). siSCR, positive; siRNA1, negative control. Blue, nuclei (DAPI); scale bar, 5  $\mu$ m (see also Supplementary Fig. 10 for uncropped images). **e**, Northern blotting detects circular but not linear CDR1as in HEK293 cells. Total, HEK293 RNA; circular, head-to-tail probe; circ+lin, probe within splice sites; IVT lin., *in vitro* transcribed, linear CDR1as RNA. **f**, Circular CDR1as is highly expressed (qPCR, error bars indicate standard deviation). **g**, CDR1as. Blue, seed matches; dark red, AGO PAR-CLIP reads; bright red, crosslinked nucleotide conversions.

**Table 1 | Summary of the validation experiments**

Sample	Validation experiment	Validation success
Human (HEK293)	Head-to-tail splicing	19 of 23
	Circularity	21 of 21
	Expression >3% vinculin	12 of 21
	Expression specificity (leukocyte specific)	5 of 7
Mouse (adult brain)	Head-to-tail splicing	3 of 3
	Circularity	3 of 3
	Expression >1% $\beta$ -actin	2 of 3
<i>C. elegans</i>	Head-to-tail splicing	15 of 20
	Circularity	13 of 13
	Expression >1% <i>elf-3.d</i>	12 of 15

Most experimentally tested circRNAs are validated.

conserved in at least one other species (Supplementary Fig. 4). Interspaced sequences were less conserved, indicating that miR-7 binding sites are probably functional (Fig. 3b). Secondary structure analysis of predicted circRNA-miRNA duplexes (Methods) showed reduced base-pairing of miR-7 beyond the seed (Fig. 3c). None of the ~1,500 miR-7 complementary sites across 32 vertebrate sequences was complementary beyond position 12 of miR-7 (only three could form an 11-nucleotide duplex) (Supplementary Table 4). Slicing by mammalian Argonaute requires complementarity of positions 10 and 11 and depends on extended complementarity beyond position 12 (ref. 36). Thus, CDR1as seems optimized to be densely bound but not sliced by miR-7.

Single-molecule imaging (Methods) revealed disperse and mostly cytoplasmic CDR1as expression (HEK293 cells), consistent with miRNA sponge function (Fig. 3d and Supplementary Table 5). CDR1as circularization was assayed by northern blotting (Fig. 3e). Nicking experiments confirmed that CDR1as circRNA can be linearized and degraded (Supplementary Fig. 5a). In RNA from HEK293 cells, circularized but no additional linear CDR1as was detected (Supplementary Fig. 5b). Circular expression levels were quantified by qPCR with divergent primers calibrated by standard curves (Supplementary Table 6). CDR1as was highly expressed (~15% to ~20% of *GAPDH* expression, Fig. 3f). Estimating *GAPDH* mRNA copy number from HEK293 RNA-seq data (~1,400 molecules per cell, data not shown) suggests that CDR1as may bind up to ~20,000 miR-7 molecules per cell (Fig. 3g).

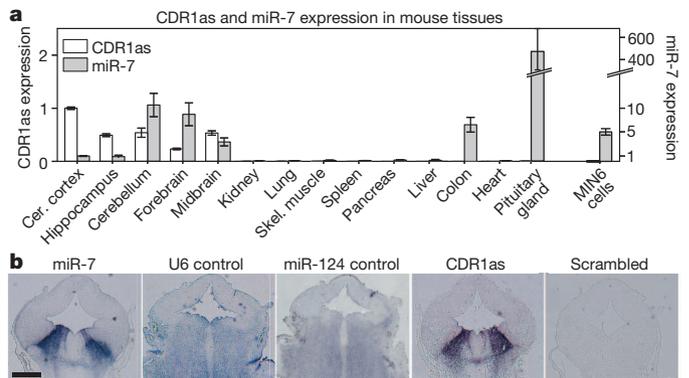
If CDR1as functions as a miR-7 sponge, its destruction could trigger downregulation of miR-7 targets. We knocked down CDR1as in HEK293 cells and monitored expression of published miR-7 targets by qPCR with externally spiked-in standards (Methods and Supplementary Fig. 5c, d). All eight miR-7 targets assayed, but also housekeeping genes, were downregulated. Nanostring technology<sup>37</sup> additionally indicated downregulation of many genes (data not shown). Furthermore, stable loss of CDR1as expression by virally delivered small hairpin RNAs led to significantly reduced migration in an *in vitro* wound closure assay (Methods, Supplementary Fig. 5e, f and Supplementary Table 7). Thus, knockdown of CDR1as affects HEK293 cells, but we could not delineate miR-7-specific effects, potentially because of indirect or miR-7-independent CDR1as function (see below).

### Co-expression of miR-7 and CDR1as in brain

If CDR1as indeed interacts with miR-7, both must be co-expressed. miR-7 is highly expressed in neuronal tissues, pancreas and pituitary gland<sup>38</sup>. Apart from HEK293 cells, a cell line probably derived from neuronal precursors in embryonic kidney<sup>39</sup>, we quantified miR-7 and CDR1as expression across mouse tissues and pancreatic-island-derived MIN6 cells (Methods and Fig. 4a). CDR1as and miR-7 were both highly expressed in brain tissues, but CDR1as was expressed at low levels or absent in non-neuronal tissues, including tissues with very high miR-7 expression. qPCR suggested that CDR1as is exclusively circular in adult and embryonic mouse brain (Supplementary Fig. 5g, h). Thus, CDR1as and miR-7 seem to interact specifically in neuronal tissues. Indeed, when assaying CDR1as and miR-7 in mouse brains by *in situ* hybridizations (Methods), we observed specific, similar, but not identical, expression patterns in the brain of mid-gestation (embryonic day 13.5 (E13.5)) embryos (Fig. 4b). Specifically, CDR1as and miR-7 were highly co-expressed in areas of the developing midbrain (mesencephalon)<sup>40,41</sup>. Thus, CDR1as is highly expressed, stable, cytoplasmic, not detectable as a linear RNA and shares expression domains with miR-7. Together with extensive miR-7 binding within CDR1as, CDR1as has hallmarks of a potent circular miR-7 sponge in neuronal tissues.

### Effects of miR-7 and CDR1as in zebrafish

It would be informative to knock out CDR1as in an animal model system. However, a knockout would also affect CDR1 protein, with



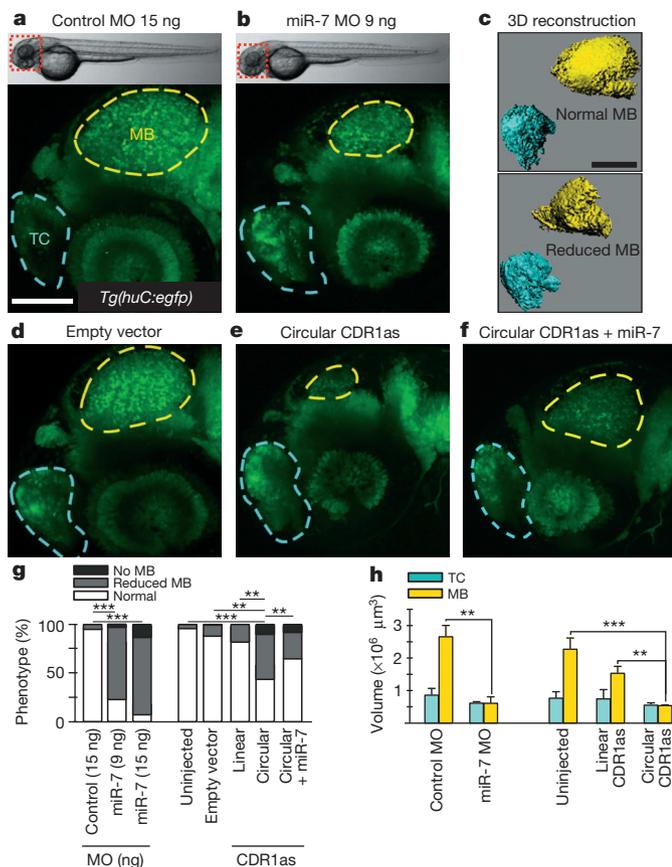
**Figure 4 | CDR1as and miR-7 have overlapping and specific expression in neuronal tissues.** **a**, Among mouse tissues and MIN6 cells (qPCR, relative to cerebral cortex expression; error bars indicate standard deviations; see Supplementary Fig. 9a for miR-122 control) neuronal tissues co-express miR-7 and CDR1as. **b**, *In situ* staining of CDR1as and miR-7 in mouse embryo brain E13.5 (U6 and miR-124, positive control; scrambled probe, negative control). Scale bar, 1 mm.

unknown consequences. This problem is circumvented when using zebrafish (*Danio rerio*) as an animal model. According to our bioinformatic analyses (not shown) zebrafish has lost the *cdr1* locus, whereas miR-7 is conserved and highly expressed in the embryonic brain<sup>42</sup>. Thus, we can test whether miR-7 has a loss-of-function phenotype and if this phenotype can be induced by introduction of mammalian CDR1as RNA. We injected morpholinos to knock down mature miR-7 expression in zebrafish embryos (Methods). At a dose of 9 ng of miR-7 morpholino, the embryos did not show overall morphological defects but reproducibly, and in two independent genetic backgrounds (Supplementary Fig. 6a–c), developed brain defects (Fig. 5a, b). In particular, ~70% showed a consistent and clear reduction in midbrain size, and an additional ~5% of animals had almost completely lost their midbrains. Of note, the telencephalon at the anterior tip of the brain was not affected in size. Brain volumes were also measured based on confocal three-dimensional stacks (Fig. 5c and Supplementary Fig. 7). Reduction of the midbrain size correlated with miR-7 inhibition in the respective animals (Supplementary Fig. 6d). These data provide evidence that miR-7 loss-of-function causes a specific reduction of midbrain size.

To test whether CDR1as can function as a miR-7 sponge *in vivo*, we injected embryos with plasmid DNA that expressed a linear version of the full-length human *CDR1as* sequence (Supplementary Fig. 6e, f) or a plasmid provided by the Kjemis laboratory that can produce circular CDR1as in human cells (Fig. 5d, e). qPCR analysis detected circular RNA in zebrafish embryos injected with the latter plasmid (Supplementary Fig. 8), which reproducibly and in independent genetic backgrounds lead to reduced midbrain sizes (Fig. 5g, h). Similarly, animals injected with *in vitro*-transcribed partial mouse CDR1as RNA, but not with RNA from the other strand, showed significant midbrain reduction (Supplementary Fig. 6g–i). Thus, the phenotype is probably caused by CDR1as RNA and not by an unspecific effect of RNA or DNA injection. These results provide evidence that human/mouse CDR1as transcripts are biologically active *in vivo* and impair brain development similarly to miR-7 inhibition. The midbrain reduction could be partially rescued by injecting miR-7 precursor (Fig. 5f, g), arguing that the biological effect of CDR1as expression is caused at least in part by interaction of CDR1as with miR-7.

### Discussion

We have shown that animal genomes express thousands of circRNAs from diverse genomic locations (for example, from coding and non-coding exons, intergenic regions or transcripts antisense to 5' and 3' UTRs) in a complex tissue-, cell-type- or developmental-stage-specific manner. We provided evidence that CDR1as can act as a



**Figure 5 | In zebrafish, knockdown of miR-7 or expression of CDR1as causes midbrain defects.** **a, b**, Neuronal reporter (*Tg(huC:egfp)*) embryos (top, light microscopy) 48 h post fertilization (bottom, representative confocal z-stack projections; blue dashed line, telencephalon (TC) (control); yellow dashed line, midbrain (MB)). Embryos after injection of 9 ng miR-7 morpholino (MO) (**b**) display a reduction in midbrain size. Panel **a** shows a representative embryo injected with 15 ng control morpholino. **c**, Three-dimensional volumetric reconstructions. **d**, Empty vector control. **e**, Expression vector encoding human circular CDR1as. **f**, Rescue experiment with miR-7 precursor. **g**, Phenotype penetrance (% of embryos, miR-7 MO,  $n = 135$ ; uninjected,  $n = 83$ ; empty vector,  $n = 91$ ; linear CDR1as,  $n = 258$ ; circular CDR1as,  $n = 153$ ; circular CDR1as plus miR-7 precursor,  $n = 217$ ). Phenotype distribution derived from at least three independent experiments. Scale bar, 0.1 mm.  $**P < 0.01$ ;  $***P < 0.001$  in Student's *t*-test for normal midbrain, reduced midbrain (see also Supplementary Fig. 6). **h**, Phenotype quantification (Methods). Error bars indicate standard deviation  $n = 3$  per group.

post-transcriptional regulator by binding miR-7 in brain tissues: (1) CDR1as is densely bound by miRNA effector molecules; (2) CDR1as harbours 74 miR-7 seed matches, often deeply conserved; (3) CDR1as is expressed highly, stably and mostly cytoplasmic; (4) CDR1as and miR-7 share specific expression domains in mouse embryonic brain; (5) human/mouse CDR1as is circularized *in vivo* and is not detectable as a linear molecule; (6) human/mouse CDR1as sequences, when injected into zebrafish, and miR-7 knock down have similar phenotypes in brain. While zebrafish circularization of human CDR1as may be incomplete, the midbrain phenotype was stronger compared to expressing linear CDR1as RNA that lacks circularization splice sites. Although the two DNA plasmids used carry identical promoters and were injected in equal concentrations, we cannot rule out the possibility that the difference in midbrain phenotype strength may be explained by other factors. However, because of the observed extreme stability of CDR1as and circRNAs in general, our data argue that circRNAs can be used as potent inhibitors of miRNAs or RBPs. Future studies should elucidate how CDR1as can be converted into a linear molecule and targeted for degradation. miR-671 can trigger destruction of

CDR1as<sup>9</sup>. Thus, CDR1as may function to transport miR-7 to subcellular locations, where miR-671 could trigger release of its cargo. Known functions of miR-7 targets such as PAK1 and FAK1 support these speculations<sup>43,44</sup>.

The phenotype induced by CDR1as expression in zebrafish was only partially rescued by expressing miR-7, indicating that CDR1as could have functions beyond sequestering miR-7. This idea is supported by *in situ* hybridization in mouse adult hippocampus (Supplementary Fig. 9b) where areas staining for CDR1as but not miR-7 were observed. What could be additional functions of circRNAs beyond acting as sponges? As a single-stranded RNA, CDR1as could, for example, bind *in trans* 3' UTRs of target mRNAs to regulate their expression. It is even possible that miR-7 binds CDR1as to silence these trans-acting activities. Alternatively, CDR1as could be involved in the assembly of larger complexes of RNA or protein, perhaps similar to other low-complexity molecules<sup>45</sup>.

How many other circRNAs exist? In this study, we identified approximately 2,000 human, 1,900 mouse and 700 nematode circRNAs from sequencing data, and our validation experiments confirmed most of the 50 tested circRNAs. However, we analysed only a few tissues/developmental stages with stringent cutoffs. Thus, the true number of circRNAs is almost certainly much larger. Although CDR1as is an extreme case, many circRNAs have conserved seed matches. For example, circRNA from the SRY locus<sup>6</sup> has seed sites for murine miRNAs. Therefore, circRNAs probably compete with other RNAs for miRNA binding. Sequence analyses indicated that coding exons serve additional, presumably regulatory functions when expressed within circRNAs, whereas intergenic or intronic circRNAs generally showed only weak conservation. Because we detected thousands of circRNAs, it is appealing to speculate that occasional circularization of exons is easy to evolve and may provide a mechanism for rapid evolution of stably and well expressed regulatory RNAs. Of note, we detected multiple seed matches for viral miRNAs within human circRNAs (not shown). However, there is no reason to think that circRNAs function predominantly to bind miRNAs. As known in bacteria, the decoy mechanism underlying miRNA sponges could be important also for RBPs<sup>46,47</sup>. Similarly, circRNAs could function to store, sort, or localize RBPs. In summary, our data suggest that circRNAs form a class of post-transcriptional regulators which compete with other RNAs for binding by miRNAs and RBPs and may generally function in modulating the local free concentration of RBPs, RNAs, or their binding sites.

*Note added in proof:* While this paper was under review, circular RNAs in fibroblasts were described<sup>51</sup>.

## METHODS SUMMARY

**Computational pipeline for predicting circRNAs from ribominus sequencing data.** A detailed description of the computational methods is given in the Methods.

**Cell culture and treatments.** HEK293, HEK293TN and HEK293 Flp-In 293 T-REx (Life Technologies) were cultured following standard protocols. Transcription was blocked by adding 2 μg ml<sup>-1</sup> actinomycin D (Sigma). RNase R (Epicentre Biotechnologies) treatment (3 U μg<sup>-1</sup>) was performed on total RNA (5 μg) at 37 °C for 15 min. qPCR primers are listed in Supplementary Table 8.

**Single-molecule RNA fluorescence *in situ* hybridization (smRNA FISH).** Stellaris Oligonucleotide probes complementary to CDR1as were designed using the Stellaris Probe Designer (Biosearch Technologies). Probe pools were obtained from BioCat GmbH as conjugates coupled to Quasar 670. Probes were hybridized at 125 nM at 37 °C. Images were acquired on an inverted Nikon Ti microscope.

**Mouse strains and *in situ* hybridization.** *In situ* hybridization (ISH) was performed on paraffin tissue sections from B6129SF1/J wild-type mice as described<sup>48</sup> using locked nucleic acid (LNA) probes or RNAs obtained by *in vitro* transcription on PCR products.

**Zebrafish methods.** *Tg(huC:egfp)* and *Tg(Xia.Tubb:dsRED)* transgenic zebrafish lines were used<sup>49,50</sup>. Morpholino antisense oligomers were injected into the yolk of single-cell-stage embryos. Furthermore, two pCS2+ plasmids coding for full-length linear CDR1as or CDR1as plus upstream and downstream sequence that can express circular CDR1as in human cells (courtesy of the Kjems laboratory)

were injected. Confocal imaging was performed using Carl Zeiss MicroImaging. Reduced midbrain development was defined as >50% smaller than the mean size of controls. Each experimental group was evaluated in at least three independent experiments; a minimum of 80 individual embryos per group was examined.

**Full Methods** and any associated references are available in the online version of the paper.

Received 11 September 2012; accepted 24 January 2013.

Published online 27 February 2013.

- Sanger, H. L., Klotz, G., Riesner, D., Gross, H. J. & Kleinschmidt, A. K. Viroids are single-stranded covalently closed circular RNA molecules existing as highly base-paired rod-like structures. *Proc. Natl Acad. Sci. USA* **73**, 3852–3856 (1976).
- Grabowski, P. J., Zaugg, A. J. & Cech, T. R. The intervening sequence of the ribosomal RNA precursor is converted to a circular RNA in isolated nuclei of Tetrahymena. *Cell* **23**, 467–476 (1981).
- Danan, M., Schwartz, S., Edelheit, S. & Sorek, R. Transcriptome-wide discovery of circular RNAs in Archaea. *Nucleic Acids Res.* **40**, 3131–3142 (2012).
- Nigro, J. M. *et al.* Scrambled exons. *Cell* **64**, 607–613 (1991).
- Cocquerelle, C., Mascrez, B., Hetuin, D. & Bailleul, B. Mis-splicing yields circular RNA molecules. *FASEB J.* **7**, 155–160 (1993).
- Capel, B. *et al.* Circular transcripts of the testis-determining gene Sry in adult mouse testis. *Cell* **73**, 1019–1030 (1993).
- Chao, C. W., Chan, D. C., Kuo, A. & Leder, P. The mouse formin (Fmn) gene: abundant circular RNA transcripts and gene-targeted deletion analysis. *Mol. Med.* **4**, 614–628 (1998).
- Burd, C. E. *et al.* Expression of linear and novel circular forms of an INK4/ARF-associated non-coding RNA correlates with atherosclerosis risk. *PLoS Genet.* **6**, e1001233 (2010).
- Hansen, T. B. *et al.* miRNA-dependent gene silencing involving Ago2-mediated cleavage of a circular antisense RNA. *EMBO J.* **30**, 4414–4422 (2011).
- Salzman, J., Gawad, C., Wang, P. L., Lacayo, N. & Brown, P. O. Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types. *PLoS ONE* **7**, e30733 (2012).
- Ambros, V. The functions of animal microRNAs. *Nature* **431**, 350–355 (2004).
- Baek, D. *et al.* The impact of microRNAs on protein output. *Nature* **455**, 64–71 (2008).
- Selbach, M. *et al.* Widespread changes in protein synthesis induced by microRNAs. *Nature* **455**, 58–63 (2008).
- Bartel, D. P. MicroRNAs: target recognition and regulatory functions. *Cell* **136**, 215–233 (2009).
- Krek, A. *et al.* Combinatorial microRNA target predictions. *Nature Genet.* **37**, 495–500 (2005).
- Lewis, B. P., Burge, C. B. & Bartel, D. P. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120**, 15–20 (2005).
- Xie, X. *et al.* Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals. *Nature* **434**, 338–345 (2005).
- Friedman, R. C., Farh, K. K., Burge, C. B. & Bartel, D. P. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.* **19**, 92–105 (2009).
- Ebert, M. S., Neilson, J. R. & Sharp, P. A. MicroRNA sponges: competitive inhibitors of small RNAs in mammalian cells. *Nature Methods* **4**, 721–726 (2007).
- Franco-Zorrilla, J. M. *et al.* Target mimicry provides a new mechanism for regulation of microRNA activity. *Nature Genet.* **39**, 1033–1037 (2007).
- Poliseno, L. *et al.* A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature* **465**, 1033–1038 (2010).
- Tay, Y. *et al.* Coding-independent regulation of the tumor suppressor PTEN by competing endogenous mRNAs. *Cell* **147**, 344–357 (2011).
- Cesana, M. *et al.* A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. *Cell* **147**, 358–369 (2011).
- Ebert, M. S. & Sharp, P. A. Emerging roles for natural microRNA sponges. *Curr. Biol.* **20**, R858–R861 (2010).
- Vivancos, A. P., Guell, M., Dohm, J. C., Serrano, L. & Himmelbauer, H. Strand-specific deep sequencing of the transcriptome. *Genome Res.* **20**, 989–999 (2010).
- Huang, R. *et al.* An RNA-Seq strategy to detect the complete coding and non-coding transcriptome including full-length imprinted macro ncRNAs. *PLoS ONE* **6**, e27288 (2011).
- Kent, W. J. *et al.* The human genome browser at UCSC. *Genome Res.* **12**, 996–1006 (2002).
- Pruitt, K. D., Tatusova, T. & Maglott, D. R. NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* **33**, D501–D504 (2005).
- Cabali, M. N. *et al.* Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* **25**, 1915–1927 (2011).
- Suzuki, H. *et al.* Characterization of RNase R-digested cellular RNA source that consists of lariat and circular RNAs from pre-mRNA splicing. *Nucleic Acids Res.* **34**, e63 (2006).
- Iwai, Y., Akahane, K., Pluznik, D. H. & Cohen, R. B. Ca<sup>2+</sup> ionophore A23187-dependent stabilization of granulocyte-macrophage colony-stimulating factor messenger RNA in murine thymoma EL-4 cells is mediated through two distinct regions in the 3'-untranslated region. *J. Immunol.* **150**, 4386–4394 (1993).
- Hafner, M. *et al.* Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141**, 129–141 (2010).
- Lebedeva, S. *et al.* Transcriptome-wide analysis of regulatory interactions of the RNA-binding protein HuR. *Mol. Cell* **43**, 340–352 (2011).
- Baltz, A. G. *et al.* The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. *Mol. Cell* **46**, 674–690 (2012).
- Dropcho, E. J., Chen, Y. T., Posner, J. B. & Old, L. J. Cloning of a brain protein identified by autoantibodies from a patient with paraneoplastic cerebellar degeneration. *Proc. Natl Acad. Sci. USA* **84**, 4552–4556 (1987).
- Wee, L. M., Flores-Jasso, C. F., Salomon, W. E. & Zamore, P. D. Argonaute divides its RNA guide into domains with distinct functions and RNA-binding properties. *Cell* **151**, 1055–1067 (2012).
- Geiss, G. K. *et al.* Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nature Biotechnol.* **26**, 317–325 (2008).
- Landgraf, P. *et al.* A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell* **129**, 1401–1414 (2007).
- Shaw, G., Morse, S., Ararat, M. & Graham, F. L. Preferential transformation of human neuronal cells by human adenoviruses and the origin of HEK 293 cells. *FASEB J.* **16**, 869–871 (2002).
- Kaufman, M. H. & Bard, J. B. L. *The Anatomical Basis of Mouse Development* (Academic, 1999).
- Schambra, U. *Prenatal Mouse Brain Atlas* (Springer, 2008).
- Kapsimali, M. *et al.* MicroRNAs show a wide diversity of expression profiles in the developing and mature central nervous system. *Genome Biol.* **8**, R173 (2007).
- Jacobs, T. *et al.* Localized activation of p21-activated kinase controls neuronal polarity and morphology. *J. Neurosci.* **27**, 8604–8615 (2007).
- Chacon, M. R. *et al.* Focal adhesion kinase regulates actin nucleation and neuronal filopodia formation during axonal growth. *Development* **139**, 3200–3210 (2012).
- Kato, M. *et al.* Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. *Cell* **149**, 753–767 (2012).
- Romeo, T. Global regulation by the small RNA-binding protein CsrA and the non-coding RNA molecule CsrB. *Mol. Microbiol.* **29**, 1321–1330 (1998).
- Gottesman, S. The small RNA regulators of *Escherichia coli*: roles and mechanisms. *Annu. Rev. Microbiol.* **58**, 303–328 (2004).
- Huelsken, J., Vogel, R., Erdmann, B., Cotsarelis, G. & Birchmeier, W.  $\beta$ -Catenin controls hair follicle morphogenesis and stem cell differentiation in the skin. *Cell* **105**, 533–545 (2001).
- Park, H. C. *et al.* Analysis of upstream elements in the HuC promoter leads to the establishment of transgenic zebrafish with fluorescent neurons. *Dev. Biol.* **227**, 279–293 (2000).
- Peri, F. & Nusslein-Volhard, C. Live imaging of neuronal degradation by microglia reveals a role for v0-ATPase a1 in phagosomal fusion *in vivo*. *Cell* **133**, 916–927 (2008).
- Jeck, W. R. *et al.* Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA* **19**, 1–17 (2013).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank M. Feldkamp and C. Langnick (laboratory of W. Chen) for Illumina sequencing runs. We thank J. Kjems for sending us a plasmid encoding circular human CDR1as for our zebrafish experiments. We thank K. Meier for technical assistance with zebrafish experiments and A. Sporbert from the confocal imaging facility. We thank A. Ivanov for assisting in bioinformatic analysis. N.R. thanks E. Westhof for useful discussions. We acknowledge the following funding sources: PhD program of the Max-Delbrück-Center (MDC) (S.M., F.T., L.H.G.); the MDC-NYU exchange program (M.M.); BMBF project 1210182, 'MiRNAs as therapeutic targets' (A.E.); DFG for KFO218 (U.Z.); Helmholtz Association for the 'MDC Systems Biology Network', MSBN (S.D.M.); BMBF support for the DZHK (F.I.N. and N.R.); Center for Stroke Research Berlin (J.K., F.I.N.). Funding for the group of M.L. is supported by BMBF-funding for the Berlin Institute for Medical Systems Biology (0315362C).

**Author Contributions** S.M., M.J., A.E. and F.T. contributed equally. S.M. performed many experiments, assisted by L.M. M.J. and A.E. carried out most of the computation, with contributions from N.R. and S.D.M. F.T. performed the circRNA validation experiments. A.R. performed all northern experiments. L.H.G. and M.M. contributed AGO PAR-CLIP experiments and HEK293 ribominus data, supervised by M.L. C.K. designed and carried out the single molecule experiments, in part together with A.L. U.Z. performed the mouse experiments. J.K. contributed the zebrafish experiments, supervised by F.I.N. N.R. designed and supervised the project. N.R. and M.J. wrote the paper.

**Author Information** Sequencing data have been deposited at GEO under accession number GSE43574. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to N.R. ([rajewsky@mdc-berlin.de](mailto:rajewsky@mdc-berlin.de)).

## METHODS

**Computational pipeline for predicting circRNAs from ribominus sequencing data.** Reference genomes (human hg19 (February 2009, GRCh37), mouse mm9 (July 2007, NB137/mm9), *C. elegans* ce6 (May 2008, WormBase v. WS190)) were downloaded from the UCSC genome browser (<http://genome.ucsc.edu/>)<sup>27</sup>. In a first step, reads that aligned contiguously and full length to the genomes were discarded. From the remaining reads we extracted 20mers from both ends and aligned them independently to find unique anchor positions within spliced exons. Anchors that aligned in the reversed orientation (head-to-tail) indicated circRNA splicing (compare main Fig. 1a). We extended the anchor alignments such that the complete read aligns and the breakpoints were flanked by GU/AG splice sites. Ambiguous breakpoints were discarded. We used the short-read mapper Bowtie 2 (ref. 52). Initially, ribominus reads were aligned in end-to-end mode to the genome:

```
$ bowtie2 -p16 --very-sensitive --phred64 --mm -M20 --score-min=C, -15, 0 -q -x <index> -U reads.qfa 2> bowtie2.log | samtools view -hbuS - | samtools sort -sample_vs_genome
```

The unmapped reads were separated and run through a custom script to split the reads as indicated in Fig. 1a to obtain 20-nucleotide anchors from both ends of the read:

```
$ samtools view -hf4 sample_vs_genome.bam | samtools view -Sb -> unmapped_sample.bam
```

```
$. /unmapped2anchors.py unmapped_sample.bam | gzip > sample_anchors.qfa.gz
```

Here is an example of two anchor pairs in the FASTQ format; the original read was kept as part of the first anchors identifier to simplify downstream analysis:

```
@s_8_1_0001_qseq_14_A_NCCCGCCTCACCGGGTCAGTGAAAAACGA
TCAGAGTAGTGGTCTTCCTCCGGCGGCCCGCGCGCCGCGCTGC
NCCCGCCTCACCGGGTCAGT
+
#BB@?@AB@; = @B; B@@58 (
@s_8_1_0001_qseq_14_B
CCCCGCGCGCGCCGCGCTGC
+
; . : ( (>) > 0 ; . 8#####
```

Next the anchors were aligned individually to the reference, keeping their paired ordering. The resulting alignments were read by another custom script that jointly evaluates consecutive anchor alignments belonging to the same original read, performs extensions of the anchor alignments, and collects statistics on splice sites. After the run completes, the script outputs all detected splice junctions (linear and circular) in a UCSC BED-like format with extra columns holding quality statistics, read counts etc. The original full-length reads that support each junction are written to stderr:

```
$ bowtie2 -p16 --reorder --mm -M20 --score-min=C, -15, 0 -q -x genome -U
sample_anchors.qfa.gz | ./find_circ.py -S hg19 -p sample_ -s sample/sites.log >
sample/sites.bed 2> sample/sites.reads
```

The resulting BED-like file is readily filtered for minimal quality cutoffs to produce the reported circRNA candidates. In particular, we demanded the following: (1) GU/AG flanking the splice sites (built in); (2) unambiguous breakpoint detection; (3) a maximum of two mismatches in the extension procedure; (4) the breakpoint cannot reside more than 2 nucleotides inside an anchor; (5) at least two independent reads (each distinct sequence only counted once per sample) support the junction; (6) unique anchor alignments with a safety margin to the next-best alignment of at least one anchor above 35 points (~more than two extra mismatches in high-quality bases); and (7) a genomic distance between the two splice sites of no more than 100 kb (only a small percentage of the data). As the ribosomal DNA cluster is part of the *C. elegans* genome assembly (ce6) and ribosomal pre-RNAs could give rise to circular RNAs by mechanisms independent

of the spliceosome, we discarded 130 candidates that mapped to the rDNA cluster on chrI:15,060,286-15,071,020.

**Permutation testing.** To test the robustness of the circRNA detection pipeline we altered the sequence of real sequencing reads in different ways at the step of anchor generation. We (1) reversed either anchor; (2) reversed the complete read; (3) randomly reassigned anchors between reads; or (4) reverse complemented the read (as a positive control). Although the reverse complement recovered the same output as expected, the various permutations led to only very few candidate predictions, well below 0.2% of the output with unpermuted reads and in excellent agreement with the results from simulated reads (Supplementary Fig. 1c).

**HEK293 RNA-seq after rRNA depletion (RibominusSeq).** Total HEK293 RNA was isolated using Trizol as recommended by the manufacturer. Ribosomal RNA was depleted from total RNA using the Ribominus kit (Invitrogen). A cDNA library was generated from rRNA-depleted RNA according to the Illumina RNA-seq protocol. The cDNA library was sequenced on an Illumina GAIIx by a 2 × 76 bp run.

***C. elegans* oocyte isolation.** Oocytes were isolated from worms carrying a temperature-sensitive (TS) allele for *fem-1* (unovulated oocytes BA17[*fem-1(hc17ts)*] strain) and *spe-9* (partially ovulated oocytes BA671[*spe-9(hc88ts)*]) as described previously<sup>53</sup>. Oocytes were washed at least four times in PBS containing protease inhibitors (Sigma-Aldrich) to separate from worm debris. Oocyte purity was observed under the dissection scope (Zeiss). Oocytes were extracted from young adults to enrich for non-endomitotic oocytes, which was also checked by fluorescence microscopy (Zeiss) with a nuclear dye. Oocytes isolated from *fem-1* or *spe-9* mutant background worms are hereafter referred to as *fem-1* oocytes and *spe-9* oocytes, respectively.

***C. elegans* sperm isolation.** Sperm was isolated in principle as described previously<sup>54</sup> from male worms obtained from a *fog-2(q71)* mutant background. Males were cut in cold PBS containing protease inhibitors (Sigma-Aldrich). Sperm was subsequently purified by filtration (3 × 40 μm nylon mesh, 2 × 10 μm nylon mesh) and a series of differential centrifugations (30 min 300g, 10 min 450g) and washed twice in cold PBS. Sperm was subsequently activated by incubation in PBS containing 200 μg ml<sup>-1</sup> Pronase (Sigma-Aldrich) for 30 min at 25 °C. Sperm purity is around 70% spermatids and spermatozoa contaminated with around 30% primary and secondary spermatocytes, as observed under oil immersion microscope.

***C. elegans* isolation of 1-cell- and 2-cell-stage embryos.** 1-cell and 2-cell-stage embryos were obtained by fluorescence-activated cell sorting as described previously<sup>55</sup>. Microscopic examination of the sorted embryos indicated that the 1-cell-stage sample was virtually pure (>98% one-cell stage embryos), whereas the 2-cell-stage embryo sample was a mixture of 1-cell-stage (40%), 2-cell-stage (55%) and older (<5%) embryos. Moreover, purity of the stages was further validated by checking for marker gene expression.

**Ribominus RNA preparation from *C. elegans* samples.** We used a kit that was developed for human and mouse samples, but still performs sufficiently to enrich mRNAs up to 30% in *C. elegans*. Most of the remaining reads mapped to ribosomal RNAs. 1 μg of total RNA per sample was depleted from rRNAs with the Ribominus Transcriptome kit (Invitrogen) according to the manufacturer's instructions with the modification that annealing of LNA probes to total RNA was performed in a thermocycler (Eppendorf) with a temperature decrease from 70 to 37 °C at a rate of 1 °C per min. Depletion of rRNAs was validated by capillary gel electrophoresis on a Bioanalyzer (Agilent). The ribominus RNA was then processed for sequencing library preparation according to the Illumina protocol.

**Cluster generation and sequencing of *C. elegans* libraries.** Cluster generation as well as sequencing of the prepared libraries was performed on the Illumina cluster station (Illumina) and sequenced on the HiSeq2000 according to the manufacturer's protocols (Illumina).

**Human gene models.** We obtained gene models for RefSeq transcripts (12 December 2011), non-coding RNAs<sup>29,56</sup> and the rnaGene and tRNA tracks from the UCSC table browser (23 April 2012)<sup>27</sup>.

**Intersection of circRNAs with known transcripts.** Our computational screen identifies only the splice sites that lead to circularization but not the internal exon/intron structure of circular RNAs. To perform analyses of the sequence content of circRNAs we therefore inferred as much as possible from annotated transcripts. The conservative assumption was that as little as possible should be spliced out. On the other hand, coincidence of circRNA splice sites with exonic boundaries inside a transcript were considered as an indicator for relevant agreement and internal introns appear to be spliced out (Supplementary Fig. 2e). We therefore sorted all overlapping transcripts hierarchically by (1) splice-site coincidence (2, 1, or 0); (2) total amount of exonic sequence between the splice sites; (3) total amount of coding sequence. The latter was used to break ties only and helped the annotation process. If one or both splice sites fell into an exon of the best matching transcript, the corresponding exon boundary was trimmed. Likewise, if it fell

into an intron or beyond transcript bounds, the closest exon was extended to match the circRNA boundaries. circRNA start/end coordinates were never altered. If no annotated exons overlapped the circRNA we assumed a single-exon circRNA. The resulting annotation of circRNAs is based on the best matching transcript and may in some cases not represent the ideal choice. Changing the annotation rules, however, did not substantially change the numbers in Fig. 1d. **Finding circRNAs conserved between human and mouse.** We reasoned that when comparing two species, the cutoff of two independent reads in each of them could be dropped, as orthologous circRNAs would automatically be supported by two independently produced reads via the intersection. We therefore mapped all mouse circRNA candidates with less stringent filtering to human genome coordinates using the UCSC liftOver tool<sup>57</sup>. The mapped mouse circRNAs were compared with independently identified human circRNAs, yielding 229 circRNAs with precisely orthologous splice sites between human and mouse. Of these, 223 were composed exclusively of coding exons and were subsequently used for our conservation analysis (Fig. 1f). When intersecting the reported sets of circRNAs supported by two independent reads in each species, we found 81 conserved circRNAs (supported by at least 4 reads in total).

**Conserved element counting.** We downloaded genome-wide human (hg19) phyloP conservation score<sup>58</sup> tracks derived from genome alignments of placental mammals from UCSC<sup>27</sup>. We interrogated the genome-wide profile inside circRNAs in two different ways. (1) Intergenic and intronic circRNAs. We read out the conservation scores along the complete circRNA and searched for blocks of at least 6-nucleotide length that exceeded a conservation score of 0.3 for intergenic and 0.5 for intronic circRNAs. The different cutoffs empirically adjust for the different background levels of conservation and were also used on the respective controls. For each circRNA, we computed the cumulative length of all such blocks and normalized it by the genomic length of the circRNA. Artefacts of constant positive conservation scores in the phyloP profile, apparently caused by missing alignment data, were removed with an entropy filter (this did not qualitatively affect the results). circRNAs annotated as intronic by the best-match procedure explained above that had any overlap with exons in alternative transcripts on either strand (five cases) were removed from the analysis. The resulting distributions are shown in Supplementary Fig. 1h, i. (2) Coding exon circRNAs. We used the best-match strategy outlined above to construct an estimated 'exon-chain' for the circRNAs that overlapped exclusively coding sequence. Using this chain we *in silico* 'spliced' out the corresponding blocks of the conservation score profile. We kept track of the frame and sorted the conservation scores into separate bins for each codon position. In addition to this, we also recorded conservation scores in the remaining pieces of coding sequence ('outside' the circRNA) as a control. However, we observed that the level of conservation is systematically different between internal parts of the coding sequence and the amino- or carboxy-terminal parts (not shown). We therefore randomly generated chains of internal exons, mimicking the exon-number distribution of real circRNAs, as a control. When analysing the circRNAs conserved between human and mouse, it became furthermore apparent that we also needed to adjust for the higher level of overall conservation. High expression generally correlates with conservation and thus, an expression cutoff was enforced on the transcripts used to generate random controls. This resulted in a good to conservative match with the actual circRNAs (Supplementary Fig. 1j, k).

**Overlap of identified circRNAs with published circular RNAs.** A number of studies in human have reported evidence for circRNAs which derive from exons of DCC<sup>4</sup>, ETS1<sup>5</sup> and a non-coding RNA from the human *INK4/ARF* locus<sup>8</sup> and the CDR1as locus<sup>9</sup>. Additionally, circRNAs from exons of the genes *CAMSAP1*, *FBXW4*, *MAN1A2*, *REXO4*, *RNF220* and *ZKSCAN1* have been recently experimentally validated<sup>10</sup>. For the four genes from the latter study, where we had ribominus data from the tissues in which these circRNAs were predicted (leukocytes), we recovered validated circRNAs from all of them (*ZKSCAN1*, *CAMSAP1*, *FBXW4*, *MAN1A2*).

**Cell culture and treatments.** HEK293 (Fig. 3f), HEK293TN (for virus production) and HEK293 Flp-In T-REx 293 (Life Technologies, all other experiments) were cultured in Dulbecco's modified Eagle medium GlutaMax (Gibco) 4.5 g l<sup>-1</sup> glucose, supplemented with 10% FCS, 20 U ml<sup>-1</sup> penicillin and streptomycin (Gibco) at 37 °C, 5% CO<sub>2</sub>. Whereas CDR1as/*GAPDH* ratios were within the given range, we observed two- to fivefold variation of CDR1as/vinculin ratios between different HEK lines. Transcription was blocked by adding 2 µg ml<sup>-1</sup> actinomycin D or DMSO as a control (Sigma-Aldrich) to the cell culture medium. For *in vitro* wound healing assays, cells were grown to confluency, the cell layer was disrupted using a 300 µl pipette tip and cells were washed once with medium. Bright-field images of cells were taken using a Axio Observer.Z1 (Zeiss) right after setting the scratch and 24 h later. The relative scratch areas were measured using ImageJ software.

**Quantitative PCR.** Total RNA from cell lines was isolated using Trizol (Invitrogen) extraction following the manufacturer's protocol. Adult B6129SF1/J mice

were dissected and tissue samples were collected directly into ice-cold Trizol for RNA preparation. *Caenorhabditis elegans* RNA was isolated from about 7,000 mixed stage worms by two rounds of freeze-thaw lysis in Trizol LS reagent (Invitrogen) according to the manufacturer's protocol. RNA was extracted from aqueous phase with phenol:chloroform (Ambion). RNA was precipitated with isopropanol and Glycobluie (Ambion) overnight at -20 °C or for 30 min at -80 °C, respectively. Reverse transcription was performed using M-MLV (Promega) or Superscript III with oligo(dT) primer (all Invitrogen) or random primer (Metabion). For assaying mRNA expression level, qRT-PCR was performed using SYBR-Green Fluorescein (Thermo Scientific, Fermentas) and a StepOnePlus PCR System (Applied Biosystems). Expression data in CDR1as knockdown experiments, transcriptional block and RNase R assays were normalized to *C. elegans* spike-in RNA. To this end 5–10% *C. elegans* total RNA was added to the respective Trizol sample and qPCR primer for *ama-1* or *eif-3.d* were used. Mouse expression data were normalized to *Actb*. miRNA expression levels were assayed using TaqMan microRNA assays (Applied Biosystems) and normalized to sno-234. Expression levels of circRNAs described in this study were measured by qPCR using divergent primers. A list of primer sequences is available in Supplementary Table 8.

**PCR amplification and Sanger sequencing.** DNA templates were PCR amplified using BioRad Mastercycler and ThermoScientific DreamTaq Green PCR Master Mix according to the manufacturer's protocol. We performed 35 cycles of PCR. PCR products were visualized after electrophoresis in 2% ethidium bromide-stained agarose gel. To confirm the PCR results, the PCR products were purified through Agencourt AMPure XP PCR purification kit. Direct PCR product Sanger sequencing was performed by LGC Genomics Ready2 Run services. Primer P1 was provided for sequencing the product for each candidate.

**Primer design.** Divergent primers were designed for each candidate (P1, P2) to anneal at the distal ends of its sequence. As negative controls we used divergent primers for *GAPDH* and *ACTB* linear transcript in HEK293 cells, and eIF-3.D in *C. elegans*. As a further negative control for divergent primers, we used genomic DNA extracted through Qiagen DNeasy Blood & Tissue kit. As positive controls, we used convergent primers for the corresponding linear transcripts or for house-keeping genes (eIF-3.D for *C. elegans*).

**RNase R treatment.** HEK293 DNase-treated total RNA (5 µg) was incubated 15 min at 37 °C with or without 3 U µg<sup>-1</sup> of RNase R (Epicentre Biotechnologies). RNA was subsequently purified by phenol-chloroform extraction, retro-transcribed through Superscript SSIII (Invitrogen) according to the manufacturer's protocol, and used in qPCR.

**RNA nicking assay.** For partial alkaline hydrolysis (nicking) 1 µg µl<sup>-1</sup> of HEK293 total RNA was incubated in 50 mM NaHCO<sub>3</sub> for 2.5 or 5 min at 90 °C or 5 min on ice for controls. After incubation the samples were immediately resuspended in denaturing RNA sample buffer and analysed on northern blots.

**Northern blotting.** Total RNA (10–20 µg) was loaded on a 1.2% agarose gel containing 1% formaldehyde and run for 2–2.5 h in MOPS buffer.

The gel was soaked in 1×TBE for 20 min and transferred to a Hybond-N<sup>+</sup> membrane (GE Healthcare) for 1 h (15 V) using a semi-dry blotting system (Bio-Rad). Membranes were dried and ultraviolet-crosslinked (at 265 nm) 1× at 200,000 µJ cm<sup>-2</sup>. Pre-hybridization was done at 42 °C for 1 h and <sup>32</sup>P-labelled oligonucleotide DNA probes were hybridized overnight. The membranes were washed briefly in 2× SSC, 0.1% SDS at room temperature and two additional times at 55 °C for 30 min, followed by two 30-min washes in 0.2× SSC, 0.1% SDS at 50–55 °C. For data collection, the membrane was exposed to a phosphorimager screen.

**Genome alignments for detecting miRNA seed complementary sites.** Multiple species alignments for the genomic intervals, corresponding to circRNAs predicted in *C. elegans* (ce6), human (hg19) and mouse (mm9), were generated via the Galaxy server at UCSC<sup>59–61</sup>. In case that a circle was overlapping with an annotated transcript, the inferred spliced sequence was used for retrieving the alignments.

The alignments included *C. elegans*, *C. briggsae* and *C. remanei* in the first case and *Homo sapiens*, *Mus musculus*, *Rattus norvegicus*, *Bos taurus* and *Canis familiaris* in the latter two.

**C. elegans human and mouse miRNAs.** Fasta files with *C. elegans*, human and mouse miRNAs were obtained from miRBase release 16 (ref. 62). Only mature miRNAs were considered for the seed analysis. According to miRBase 16 a mature miRNA is the predominant miRNA between the two species arising from the two arms of the precursor hairpin (information that is not included in more recent versions). The miRNAs were grouped into families that share a common seed (nucleotides 2–7). There are 117, 751 and 723 miRNA families for *C. elegans*, human and mouse, respectively.

**Detecting putative miRNA seed matches.** The *C. elegans*, human and mouse multiple species alignments were scanned for putative conserved miRNA target sites for each of the mature miRNA families. A putative target site of a miRNA is a

6-nucleotide-long sequence in the genome that is the reverse complement of nucleotides 2–7 of the mature miRNA sequence. A putative target site is called conserved if it is found in *C. elegans*, *C. briggsae* and *C. remanei* in the first case or in human, mouse, rat, cow and dog in the latter.

**AGO PAR-CLIP.** Generation and growth conditions of human embryonic kidney (HEK) 293 cells and HEK293 stably expressing Flag/HA-AGO1 and Flag/HA-AGO2 were reported previously<sup>63</sup>. Stably transfected and parental HEK293 cells were labelled with 100  $\mu$ M 4-thiouridine for 16 h. After labelling, procedure followed the PAR-CLIP protocol as described<sup>32</sup>. Briefly, ultraviolet-irradiated cells were lysed in NP-40 lysis buffer. Immunoprecipitation was carried out with protein G magnetic beads (Invitrogen) coupled to anti-Flag antibody (Sigma) and to anti-AGO2 antibody<sup>64</sup> from extracts of stably transfected and parental HEK293 cells, respectively, for 1 h at 4 °C. Beads were treated with calf intestinal phosphatase (NEB) and radioactively end-labelled by T4 polynucleotide kinase (Fermentas). The crosslinked protein–RNA complexes were resolved on 4–12% NuPAGE gel (Invitrogen), and a labelled protein–RNA complex of close to 100 kDa was excised. The protein–RNA was isolated by electroelution. RNA was isolated by proteinase K treatment and phenol-chloroform extraction, reverse transcribed and PCR-amplified. The amplified cDNA was sequenced on a GAIIX (Illumina) with 36 cycles.

**Human Argonaute PAR-CLIP analysis.** We obtained Argonaute PAR-CLIP reads from ref. 32. We additionally produced 4 PAR-CLIP libraries. In total, we analysed the following PAR-CLIP data sets: AGO1\_4su\_1 (SRR048973), AGO3\_4su\_1 (SRR048976) from ref. 32; AGO1\_4su\_ML\_MM\_6, AGO2\_4su\_ML\_MM\_7, AGO2\_4su\_ML\_MM\_8, and AGO2\_4su\_3\_ML\_LG (our own data, published under GEO accession GSE43574).

Redundant reads were collapsed (such that each distinct read sequence appears only once), aligned to the human genome (assembly hg19) using bwa 0.6.1-r104 (ref. 65), and analysed by our in-house PAR-CLIP analysis pipeline (Jens, M. *et al.*, unpublished), essentially as described in ref. 33. Briefly, reads uniquely aligning to the genome are grouped into clusters contiguously covering the reference, assigning each cluster a number of quality scores (T conversions, number of independent reads, etc.). Clusters with less than 3 reads from 3 of 6 independent AGO PAR-CLIP libraries or lacking T conversions were discarded. Remaining clusters are annotated against a comprehensive list of transcript models (see below) and collected into ‘only sense’, ‘only antisense’ and ‘intergenic/overlapping transcription’ categories based on their annotation. As PAR-CLIP sequencing preserves the directionality of RNA fragments we assume ‘only antisense’ clusters to predominantly represent false positives due to mapping artefacts (PAR-CLIP RNA is mutated and fragments are often short), and choose quality cutoffs for all clusters such that the fraction of kept ‘only antisense’ clusters is reduced to below 5%. Remaining ‘only antisense’ clusters were discarded. For Fig. 3a, uniquely aligning, collapsed reads are shown.

**AGO binding sites in *C. elegans*.** Sequencing reads from the Zisoulis Alg-1 HITS-Clip data were obtained from [http://yeolab.ucsd.edu/yeolab/Papers\\_files/ALG1\\_MT-WT\\_raw.tar.gz](http://yeolab.ucsd.edu/yeolab/Papers_files/ALG1_MT-WT_raw.tar.gz) (ref. 66). The raw sequencing data of the wild-type Alg-1 HITS-CLIP was pre-processed and mapped with the mapper module from miRDeep2 (ref. 74). The pre-processed reads were mapped with bowtie version 0.12.7 (ref. 67) to the *C. elegans* genome (ce6). All reads that overlapped when mapped to the genome were merged into bigger regions (islands). Read counts were averaged. This resulted in 24,910 islands in the *C. elegans* genome.

**Analysis of sequence conservation in CDR1as.** Genome alignments of 32 vertebrates were downloaded from the UCSC database (hg19)<sup>27</sup> and analysed for the *CDR1as* locus. Primate species other than human were discarded to not bias the analyses. The one species (cow) with more than 50% gaps in the *CDR1as* locus was also discarded. The alignments for the seed regions were then corrected. Specifically, bases that would clearly align with the seed but had been separated in the alignment by runs of gaps were re-aligned. These corrections were necessary in less than 1% of all seed sites.

For an in-depth analysis we BLATed<sup>68</sup> the human *CDR1as* sequence with 20-nucleotide flanking region against all vertebrate genomes in the UCSC genome browser and kept only hits that in turn aligned best to the human locus. The resulting sequences were used to build a multiple species alignment with MUSCLE<sup>69</sup>. The same corrections were applied as described above. This alignment was also used for Supplementary Fig. 4. Entropy was calculated in log<sub>2</sub> units and averaged across all alignment columns bracketing each human seed site by maximally 8 nucleotides.

**Analysis of miR-7 base-pairing within CDR1as.** RNAfold<sup>70</sup> was used to co-fold miR-7 with each of the 74 binding regions within CDR1as defined as the miR-7 seed match TCTTCC and the next 16 bases upstream.

**Single-molecule RNA fluorescence *in situ* hybridization (smRNA FISH).** 48 oligonucleotide probes (20 nucleotides length; spacing 2 nucleotides) complementary to the *CDR1as* transcript were designed using the Stellaris Probe

Designer version 2.0 (Biosearch Technologies) with a masking level of 4 on the human genome to achieve high probe specificity (Supplementary Table 8). Stellaris probe pools were obtained from BioCat GmbH as conjugates coupled to Quasar 670 (a Cy 5 replacement). Flp-In T-REX 293 cells (Life Technologies) were grown exponentially and seeded into LabTek 4-well chambered coverslips (1 to 2 × 10<sup>5</sup> cells per well). Hybridizations were performed according to the manufacturer’s instructions with 50 ng ml<sup>-1</sup> DAPI as nuclear counterstain; Stellaris probes were hybridized at 125 nM concentration with a stringency of 10% formamide in overnight hybridizations at 37 °C. Images were acquired on an inverted Nikon Ti microscope with a Hamatsu ORCA R2 CCD camera, a 60× NA 1.4 oil objective and Nikon NIS-Elements Ar software (version 4), using an exposure time of 50 ms for DAPI and 1–1.5 s for Quasar 670. Groups of cells for imaging were chosen in the DAPI channel; Z-stacks were acquired in the Quasar 670 channel using 0.3  $\mu$ m spacing and comprised a total depth of 6.5  $\mu$ m (5  $\mu$ m below and 1.5  $\mu$ m above the middle of the nucleus) and merged using maximum intensity.

**Mouse strains and *in situ* hybridizations.** All mice were bred and maintained in the animal facility of the Max Delbrück Centrum under specific pathogen-free conditions, in plastic cages with regular chow and water ad libitum. All aspects of animal care and experimental protocols were approved by the Berlin Animal Review Board (REG 0441/09). B6129SF1/J wild-type adult, newborns (postnatal day 1) or pregnant females (plug detection at day 0.5; embryo collection at day 13.5) were used, as indicated for each experiment, to obtain the tissues needed for RNA analysis and *in situ* hybridizations (ISH). After death, embryos or tissues were immediately frozen in liquid nitrogen and stored at –70 °C, or fixed for ISH.

Mouse brain structures were collected and named according to the anatomical guidelines of the Gene Expression Nervous System Atlas of the Rockefeller University (<http://www.gensat.org>) and the Mouse Brain Atlas ([http://www.mbl.org/mbl\\_main/atlas.html](http://www.mbl.org/mbl_main/atlas.html)).

For the RNA analysis and to clone CDR1as-specific RNA probes, two adult 1-year-old mice of both sexes were dissected, total RNA prepared and analysed. If embryos or newborns were sectioned, a minimum of two specimens were evaluated; in some instances up to 5 specimens were used.

For ISH, samples were fixed in formalin (1×PBS; 4% formaldehyde) for 12 h and post-fixed (70% ethanol, 18 h) before dehydrating and paraffin-embedding. Next, the organs were perfused with a standard protocol using a Shandon XP Hypercentre. For ISH mouse embryos or organs were cut in RNase-free conditions at 6  $\mu$ m and ISH was performed as described<sup>48</sup> with digoxigenin (DIG)-labelled RNA probes. All DIG–RNA probes were hybridized at 58 °C overnight. A total of 600 ng of the labelled probes was used per slide.

To amplify *Cdr1* sense and antisense sequences for ISH probe preparation a standard PCR-amplification was performed using mouse cerebellum cDNA. Three *Cdr1as* amplicons were generated, two of which probes are meant for the detection of both linear and circular forms using mmuCdr1\_1f 5′-TGCCAGTACCAAGGTCTTCC-3′ and mmuCdr1\_1r 5′-TTTTCTGCTGGAAGATGTCAA-3′, as well as mmuCdr1\_2f 5′-CCAGACAATCGTGATCTTCC-3′ and mmuCdr1\_2r 5′-ATCTTGCTGGAAGACTTGG-3′. In addition a probe was generated, specific to the circular probe, using the divergent primers mmuCdr1\_as\_7f 5′-CCACATCTTCCAGCATCTTT-3′ and mmuCdr1\_as\_7r 5′-TGGATCCCTTGAAGACAAA-3′ (CDR1\_as head to tail probe). All ensuing fragments were subcloned into pCR-BluntII-TOPO (Invitrogen) and verified by sequencing. Linearized plasmids were amenable for *in vitro* transcription using the T7 (antisense) or SP6 (sense) polymerase and a DIG-label nucleotide mixture according to manufacturer’s instruction (Roche Applied Science).

LNA ISHs were performed according to a protocol suggested by the manufacturer (Exiqon) with minor modifications. For individual LNAs, specific protocols were run at 51 °C (miR-7; 38485-15) or 58 °C (miR-124; 88066-15) on an InSituPro VS robot (Intavis). A pre-hybridization step was added, which consisted of an incubation of the slides at 15 °C lower than the hybridization temperature for 30 min using hybridization buffer. The antibody-blocking step was performed in the presence of 1% mouse blocking reagent (Roche 11096176001) and 10% sheep serum. The LNA probes were used at the following concentrations: miR-7 40 nM; miR-124 20 nM; U6 snRNA 1 nM; scrambled 40 nM, as suggested by miRCURY LNA microRNA ISH Optimization kit (Exiqon; 90004). Before detection all slides were washed 4× in NTMT including 1 mM Levamisole. The doubly DIG-labelled LNAs were detected by the alkaline phosphatase using the substrate BM-purple (Roche; 11442074001) at 37 °C.

**siRNA- and shRNA-mediated knock down.** CDR1as was knocked down using custom designed siRNA oligonucleotides (Sigma) and Lipofectamine RNAiMax (Invitrogen). 2 × 10<sup>6</sup> HEK293 cells were transfected with 10 nM siRNA duplex following the manufacturer’s protocol. After 12–16 h cells were harvested and subjected to RNA analysis. For stable knock down of CDR1as, 293TN cells were co-transfected with the packaging plasmids pLP1, pLP2 and the VSV-G plasmid

(Invitrogen) and pSicoR constructs<sup>71</sup> (sequences available in the Supplementary Table 8) by calcium phosphate transfections. Viral supernatants were harvested after 24 h and 48 h post transfection and filtered through a 0.44  $\mu\text{m}$  filter. For infection the viral supernatants supplemented with fresh medium and 6  $\mu\text{g ml}^{-1}$  polybrene was added to target cells. After overnight infection cells were allowed to recover for 12 h and subjected to a second round of infection. Cells were collected 48–72 h after the first infection. The list of siRNA oligonucleotides is provided in Supplementary Table 8.

**Zebrafish methods.** Zebrafish and their embryos were handled according to standard protocols<sup>72</sup> and in accordance with Max Delbrück Centrum institutional ethical guidelines. The *Tg(huC:egfp)* and the *Tg(Xia.Tubb:dsRED)* transgenic zebrafish lines have been described elsewhere<sup>49,50</sup>. Morpholino antisense oligomers (Gene Tools) were prepared at a stock concentration of 1 mM according to the manufacturer's protocol. Sequences: control morpholino, 5'-CTC TTACCTCAGTTACAATTTATA-3' (control morpholino) and morpholino targeting miR-7, 5'-ACAACAAAATCACAAGTCTCCACA-3' (miR-7 morpholino). For titration experiments we used 15 ng of control morpholino and 9 and 15 ng of miR-7 morpholino; for all other experiments we used 9 ng miR-7 morpholino. 3 nl of morpholinos were injected into the yolk of single-cell-stage embryos.

A 673-nucleotide mouse *Cdr1as* fragment was amplified from mouse cerebellar cDNA and the amplicon was subcloned into a pCR-Blunt II Topo vector (Invitrogen). The vector was linearized with KpnI or ApaI (Fermentas) *in vitro* transcribed (IVT) using T7 and SP6 RNA polymerases (Promega) and the resulting *Cdr1as* and reverse complement *Cdr1as\_control* products were used for injections (1.5 nl of 100 ng nl<sup>-1</sup>) into the cell of single-cell-stage embryos. In a repetition of these experiments the *Cdr1as* fragment amplicon was directly used as a template for IVT by exploiting T7-promoter extended forward and reverse primer.

Approximately 1.5 nl of a 50 ng  $\mu\text{l}^{-1}$  construct (backbone pCS2+) expressing the human linear or the human circular CDR1as was injected into the cell of single-cell-stage embryos (provided by the Kjems laboratory). For rescue experiments the construct containing the circular CDR1as was injected together with 1.5 nl pre-miR-7 precursor (7  $\mu\text{M}$ , pre-miR miRNA precursor ID PM10047 from Applied Biosystems). The negative control was the vector pCS2+ without insert (empty vector, 50 ng  $\mu\text{l}^{-1}$ ).

Confocal imaging was performed using a Zeiss LSM 510 microscope (Carl Zeiss MicroImaging) equipped with a 25 $\times$  objective (NA = 0.8). Embryos were anaesthetized using 0.1% tricaine and mounted in 1% agarose as described<sup>73</sup>. Confocal stacks were acquired of the brain region and processed using Zeiss ZEN software. Midbrain and telencephalon volumes were calculated using Imaris 64 $\times$ 7.6.1 software based on high-resolution three-dimensional stacks obtained from *Tg(Xia.Tubb:dsRED)* embryos. Reduced midbrain development was defined as >50% smaller than the mean size of controls.

Each experimental group was evaluated in at least three independent experiments; a minimum of 80 individual embryos per group were examined. Data are expressed as mean  $\pm$  standard deviation. Statistical analysis was performed using Student's *t*-test, and a  $P < 0.05$  was considered statistically significant.

Expression of miR-7 in zebrafish embryos at 48 hours post fertilization was normalized to expression of  $\beta$ -actin. In the miR-7 morpholino group, only embryos with a midbrain phenotype were used for the RNA expression analysis.

dre  $\beta$ -actin forward primer, 5'-TGCTGTTTTCCCCTCCATTG-3'; reverse primer, 5'-TTCTGTCCCATGCCAACCA-3'; probe sequence FAM-5'-TGGAC GACCAGACATCAGGGAGTG-3'-TAMRA.

For measuring the expression of dre-miR-7a/b we used Applied Biosystems TaqMan miR assays (ID000268, ID001088).

52. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**, 357–359 (2012).
53. Aroian, R. V., Field, C., Pruliere, G., Kenyon, C. & Alberts, B. M. Isolation of actin-associated proteins from *Caenorhabditis elegans* oocytes and their localization in the early embryo. *EMBO J.* **16**, 1541–1549 (1997).
54. L'Hernault, S. W. & Roberts, T. M. Cell biology of nematode sperm. *Methods Cell Biol.* **48**, 273–301 (1995).
55. Stoeckius, M. *et al.* Large-scale sorting of *C. elegans* embryos reveals the dynamics of small RNA expression. *Nature Methods* **6**, 745–751 (2009).
56. Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnol.* **28**, 511–515 (2010).
57. Hinrichs, A. S. *et al.* The UCSC Genome Browser Database: update 2006. *Nucleic Acids Res.* **34**, D590–D598 (2006).
58. Pollard, K. S. *et al.* Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.* **20**, 110–121 (2010).
59. Blankenberg, D. *et al.* Galaxy: a web-based genome analysis tool for experimentalists. *Curr. Protoc. Mol. Biol.* Ch. 19, Unit 19 10 11–21 (2010).
60. Giardine, B. *et al.* Galaxy: a platform for interactive large-scale genome analysis. *Genome Res.* **15**, 1451–1455 (2005).
61. Goecks, J., Nekrutenko, A., Taylor, J. & Galaxy, T. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol.* **11**, R86 (2010).
62. Griffiths-Jones, S. The microRNA Registry. *Nucleic Acids Res.* **32**, D109–D111 (2004).
63. Landthaler, M. *et al.* Molecular characterization of human Argonaute-containing ribonucleoprotein complexes and their bound target mRNAs. *RNA* **14**, 2580–2596 (2008).
64. Rudel, S., Flatley, A., Weinmann, L., Kremmer, E. & Meister, G. A multifunctional human Argonaute2-specific monoclonal antibody. *RNA* **14**, 1244–1253 (2008).
65. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
66. Zisoulis, D. G. *et al.* Comprehensive discovery of endogenous Argonaute binding sites in *Caenorhabditis elegans*. *Nature Struct. Mol. Biol.* **17**, 173–179 (2010).
67. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
68. Kent, W. J. BLAT—the BLAST-like alignment tool. *Genome Res.* **12**, 656–664 (2002).
69. Goujon, M. *et al.* A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Res.* **38**, W695–W699 (2010).
70. Bernhart, S. H. *et al.* Partition function and base pairing probabilities of RNA heterodimers. *Algorithms Mol. Biol.* **1**, 3 (2006).
71. Ventura, A. *et al.* Cre-lox-regulated conditional RNA interference from transgenes. *Proc. Natl Acad. Sci. USA* **101**, 10380–10385 (2004).
72. Westerfield, M. *The Zebrafish Book: A Guide for the Laboratory Use of Zebrafish (Brachydanio rerio)* 2nd edn (Univ. Oregon Press, 1993).
73. Krueger, J. *et al.* Flt1 acts as a negative regulator of tip cell formation and branching morphogenesis in the zebrafish embryo. *Development* **138**, 2111–2120 (2011).
74. Friedländer, M. R., Mackowiak, S. D., Li, N., Chen, W. & Rajewsky, N. miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Res.* **40**, 37–52 (2012).