PROTEIN-PROTEIN INTERACTIONS & GENOME ENGINEERING

Novel strategies to study cell polarity

Selma Waaijers

ISBN: 978-90-393-6183-2

Picture cover by: Bas van de Riet

Cover design by: Vincent Portegijs & Selma Waaijers

Layout by: Selma Waaijers

Printed by: Gildeprint

Protein-protein interactions and genome engineering

Novel strategies to study cell polarity

Eiwit-eiwit interacties en genoommodificatie

Nieuwe strategieën voor het bestuderen van celpolariteit

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de rector magnificus, prof.dr. G.J. van der Zwaan, ingevolge het besluit van het college voor promoties in het openbaar te verdedigen op vrijdag 19 september 2014 des middags te 12.45 uur

door

Selma Waaijers

geboren op 11 maart 1985
te Hilversum

Promotor: Prof.dr. S.J.L. van den Heuvel

Co-promotor: Dr. M. Boxem

CONTENTS

General introduction

Cortical polarity regulators and polarity establishment in the nematode *Caenorhabditis elegans*

Selma Waaijers

Utrecht University, Department of Biology, Padualaan 8, 3584 CH Utrecht, The Netherlands.

CONTROL OF CELL POLARITY BY EVOLUTIONARILY CONSERVED CORTICAL POLARITY REGULATORS

A polarized cell is a cell in which contents, shape, or functions are asymmetrically distributed along a polarity axis. Cell polarity guides a wide variety of cellular processes, from cell migration and intracellular transport to positioning the division plane during cell division. The importance of cell polarity is reflected by its critical role in all multicellular and many unicellular life forms. Loss of cell polarity is associated with several diseases and is one of the hallmarks of cancer.

Three highly conserved groups of proteins play a key role in regulating cell polarity. The PAR proteins are involved in polarity regulation of many cell types. The two other groups of polarity regulators, the Crumbs and Scribble groups, have more restricted functions and establish polarity in epithelia together with the PAR proteins. These three protein groups localize to the cortex of epithelial cells where they form two cortical domains: an apical and a basolateral cortical domain, with cell-cell junctions located at the interface of these two cortical domains. The localization of polarity and junctional proteins are interdependent and together they regulate the establishment and maintenance of cell polarity as well as cell-cell interactions.

The PAR proteins

The PAR loci were discovered in a screen for mutants defective in the partitioning of cell-fate specification in the one-cell *Caenorhabditis elegans* embryo.[1] Mutations in six genes, *par-1* to *par-6*, caused the first cell division to be symmetric instead of asymmetric.[1,2] During polarization of the fertilized egg, the serine-threonine kinase PAR-1 and the RING-finger-containing protein PAR-2 become localized to the posterior cortex, while the PDZ-domain-containing proteins PAR-3 and PAR-6 become localized to the anterior cortex. The serine-threonine kinase PAR-4 and the 14-3-3 protein PAR-5 are uniformly distributed during polarization of the zygote.[3,4] The atypical protein kinase C (aPKC) homologue PKC-3 and the RHO GTPase CDC-42 were identified later as core members of the PAR-3/PAR-6 complex.[5–9] With the exception of PAR-2, all PAR proteins are conserved in other organisms, where they are involved in the polarization of many different cell types.[10–12] For example, CDC42, PAR3, PAR6, and aPKC control the formation of the apical domain in epithelia, position the nucleus and centrosomes in *C. elegans* intestinal cells and control multiple aspects of polarity in migrating neurons.[13–15] PAR1 meanwhile is an important regulator of basolateral identity in epithelial cells. The cytoplasmically located 14-3-3 protein PAR5 binds phosphorylated proteins, including PAR1, PAR3, and LGL, and appears to play a role in restricting their location to the correct destination. Finally, the activity of the mammalian homolog of PAR-4, termed LKB1, is sufficient to induce the formation of microvilli in intestinal epithelial cells.[16]

The PAR3, PAR6, aPKC, and CDC42 proteins can form multiple interactions which are well described (Fig. 1). PAR6 can interact directly with all other members of the complex. PAR6 interacts through its semi-CRIB and PDZ domain with CDC42.[6,17,18] PAR6 and aPKC interact through their PB1 domains, and the PDZ domain of PAR6 can bind to the PDZ1 domain

of PAR3.[17–21] Finally, PAR3 can bind to both PAR6, through its PDZ1 domain, and aPKC, through a conserved binding site in its CR3 region.[5,22,23] Although these four proteins can form a single complex, only PAR6 and aPKC appear to act together in all situations. CDC42 is an important regulator of PAR6/aPKC, but has many other cellular targets. PAR3 can be associated with PAR6/aPKC, as in the *C. elegans* embryo, but is also frequently found at a distinct localization. For example, in *Drosophila melanogaster* epithelial cells, PAR6/aPKC localize apically while the PAR3 homolog Bazooka (Baz) is concentrated just apically of cell junctions. Relocalization of PAR3 to this subapical region involves phosphorylation of a conserved serine residue to disrupt binding to aPKC, and binding of members of the Crumbs complex to PAR6 to disrupt the PAR3/PAR6 interaction.[24–27] In *C. elegans* embryonic intestinal cells, PAR-3 localizes subapically, while PAR-6 and PKC-3 localize apically.[28]

The Crumbs complex

The Crumbs complex consists of four core members: the transmembrane protein Crumbs, the membrane-associated guanylate kinases (MAGUK) protein Stardust (Sdt), and the PDZ- and Lin-2/Lin-7 (L27)-domain-containing proteins PATJ and Lin-7.[29–32] Crumbs and Stardust were first identified in *Drosophila* as regulators of epithelial polarity in the embryo, and mutants lacking either of these two genes die of epithelial polarity defects.[29,30] The phenotypes and localization pattern of Crumbs and Sdt are nearly identical. Both Crumbs and Sdt localize to the apical membrane of epithelia, and concentrate just apical of the apical junctions.[30,33,34] Mutations in either gene can lead to loss of the apical cortical domain, whereas overexpression of Crumbs results in expansion of the apical cortical domain.[35] Thus, Crumbs and Sdt control the specification of the apical membrane domain. In addition, Crumbs is involved in the assembly of apical junctions.[30,34–37] The two additional complex members, PATJ and Lin-7, were discovered later.[31,32] The Crumbs proteins form a complex via direct protein-protein interactions (Fig. 1). Crumbs binds to Stardust via its intracellular PDZ-binding domain and Stardust binds to the L27 domains of Lin-7 and PATJ.[38–44] While Crumbs and Stardust are essential genes in *Drosophila*, mutants lacking Lin-7 or PATJ are viable.[32,45] Thus, while these four proteins are considered to form the core Crumbs complex, their *in vivo* requirements differ.

The members of the Crumbs complex are highly conserved in vertebrates. Crumbs itself is represented in vertebrates by three distinct family members: CRB1, CRB2, and CRB3. Of the three, CRB1 and CRB2 resemble *Drosophila* Crumbs in size and structure, while CRB3 is a much smaller family member, with a conserved intracellular domain but a very short extracellular domain (59 amino acids in human) lacking the EGF repeats and Laminin G-like domains present in CRB1 and CRB2. Interestingly, CRB3 is widely expressed in epithelial tissues and skeletal muscle, while CRB1 and CRB2 are only expressed in the retina and brain, with additional expression observed for CRB2 in the kidney.[46–49] In *C. elegans*, two Crumbs proteins have been described: CRB-1 and EAT-20.[50–53] In chapter 5 I describe the identification and characterization of a third *C. elegans* Crumbs homolog that resembles mammalian CRB3.

The mammalian homolog of Stardust is called protein associated with Lin-7 (PALS1, also known as MPP5). *In vitro* studies show a role for PALS1 in tight junction formation, in the recruitment of ezrin to the apical membrane of intestinal cells, in the localization of CRB1 in glial cells, and in the polarization of the myelin sheath.[54–57] PALS1 appears to play an important role in retinal epithelial cells, as flies, zebrafish and mice lacking PALS1 all display defects in the organization of the retinal epithelium.[27,58–62] In *C. elegans* there are three candidate Sdt homologs: *magu-1*, *magu-2*, and *magu-3*.[38,63] The roles of these proteins in cell polarity in *C. elegans* have not been investigated.

LIN-7 was originally identified in *C. elegans* as a cell junction-associated protein required for the basal localization of the LET-23 receptor tyrosine kinase.[64] LIN-7 contains an N-terminal L27 domain, with which it binds to Stardust/PALS1, and a C-terminal PDZ domain. In *Drosophila*, loss of DLin-7 does not lead to clear defects in epithelial polarity. In flies lacking DLin-7, photoreceptors undergo progressive, light dependent degeneration of photoreceptors, a phenotype shared with flies lacking Crumbs.[32] In mammals, three homologs of Lin7 exist (Lin7a-c). In cultured MDCK epithelial cells, knockdown of Lin7c results in defects in tight junction formation and loss of PATJ and PALS1 expression.[65] Mice lacking Lin7c show defects in epithelial polarity in the kidney, and mice lacking all three Lin7 homologs fail to localize PALS1 and PATJ to the apical side of neuronal precursor cells (NPCs).[66,67]

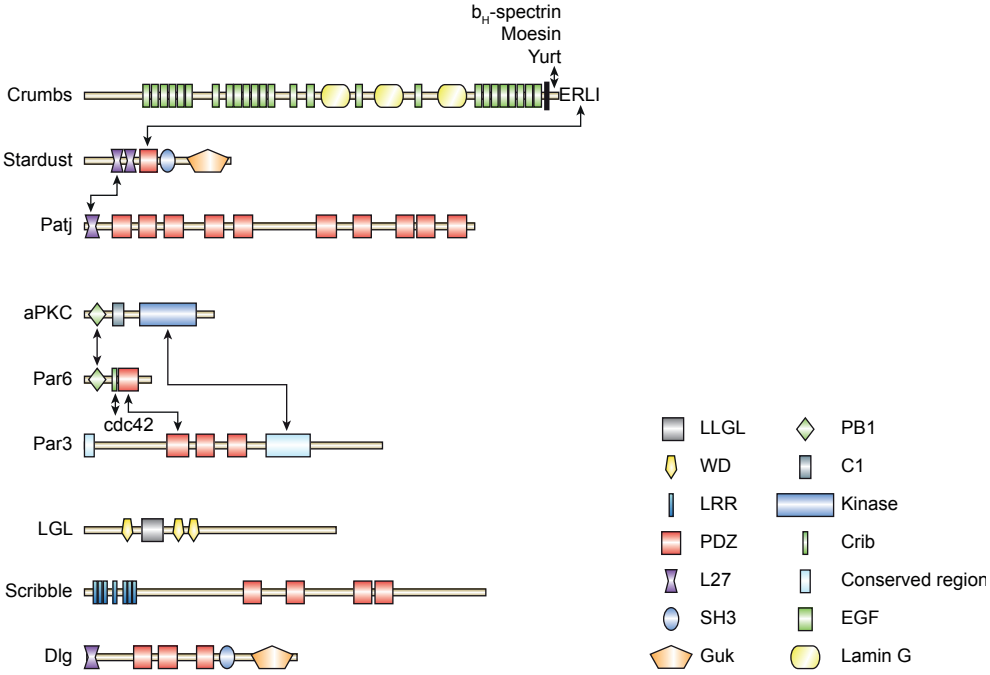


Figure 1. Schematic representation of the protein domain structure of the key polarity regulators. Arrows indicate known interactions.

The Scribble group

The members of this group, Scribble, Lethal giant larvae (Lgl), and Discs large (Dlg) were initially discovered in *Drosophila* as genes whose mutation results in loss of apical-basal polarity and neoplastic overgrowth of imaginal disc epithelia.[68,69] All three proteins are expressed in embryonic, follicular, as well as imaginal disc epithelia, and localize to the lateral membrane. Loss of any of the three genes results in overproliferation of imaginal discs, lateral localization of apical proteins, and disrupted formation of cell-cell junctions.[68,69] Finally, all three proteins are dependent on each other for their proper lateral localization. Although sometimes referred to as a complex, clear evidence for direct interactions between them is lacking.

Each of these three proteins is conserved across species. In *C. elegans*, single homologs exist for all of the three proteins. In mammals, Scribble and Lgl are represented by single genes, but analysis of DLG is complicated by the presence of at least four *DLG* genes (*DLG1-4*). Functional conservation is indicated by the finding that rat DLG1 and DLG3 can rescue the phenotype of *Drosophila dlg1* mutants.[70] However, distinct subcellular localization patterns also indicate a diversification of their roles.[71] In addition to roles in epithelial polarity, DLG4, also known as postsynaptic density protein of 95 kDa (PSD95), also fulfills important roles in synaptic development in the nervous system.[72]

As these proteins were identified as tumor suppressor in *Drosophila*, much attention has gone into identifying a similar role in mammals. The most concrete evidence for a role as tumor suppressors is the finding that Scribble and DLG proteins are targeted by viral oncoproteins.[73] E6 oncoproteins from high risk human papillomavirus (HPV) types interact with several PDZ domain containing proteins, including Scribble, DLG1, and DLG4. DLG1 and DLG4 are also targeted by human T cell leukemia virus type 1 Tax protein, and DLG1 is targeted by E43-ORF1 of the adenovirus type 9.[71] Importantly, all three viral oncoproteins depend on the presence of an intact PDZ domain for viral transformation.

Polarity is established by mutual exclusion

The PAR, Crumbs, and Scribble polarity regulators together establish two distinct membrane domains. How do these polarity regulators establish and maintain stable polarized domains? Mutual exclusion appears to be a key mechanism that drives polarization. Although the molecular details are not fully understood, phosphorylation by PAR-1 and aPKC play a major role. A number of aPKC phosphorylation targets are known, including LGL, PAR-2, and PAR3, while PAR1 is known to phosphorylate PAR3. Phosphorylation can prevent localization of the target protein through multiple mechanisms. For example, phosphorylation of LGL by aPKC may trigger a closed conformation that is unable to bind to the cortex.[74] Phosphorylation of *Drosophila* Baz by PAR1 on two serine residues results in binding of the 14-3-3 protein PAR5, which prevents Baz oligomerization as well as binding to aPKC.[75] An important question is how these mechanisms act in different contexts and ultimately result in a wide range of polarized cell types.

CELL POLARITY IN *C. ELEGANS*

Polarity in the one-cell embryo

One of the best studied examples of polarity establishment and maintenance is the polarization of the *C. elegans* one-cell embryo into a cell with an anterior and posterior cortical domain. The anterior side is defined by PAR-3/PAR-6/PKC-3/CDC-42.[5,22,76,77] PAR-1/PAR-2/LGL-1 define the posterior cortical domain.[78–81] Initially, PAR-3, PAR-6, and the atypical kinase C PKC-3 localize uniformly to the cortex, while PAR-1 and PAR-2 are cytoplasmic at that stage.[5,22,82,83] PKC-3 phosphorylates PAR-2, preventing it from associating with the cortex. The position of the paternal pronucleus and its associated centrosome after meiosis define the future posterior side of the embryo.[84] Two partially redundant mechanisms break the symmetric distribution of the PAR proteins. A network of actomyosin is present under the plasma membrane and physical interaction between the paternal centrosome and the posterior cortex causes the network to destabilize and retract from the posterior side, a phenomenon termed cortical flow. GFP::PAR-6 foci move anteriorly at the same speed as the myosin foci, suggesting that the anterior PAR proteins are displaced to the anterior side by cortical flow.[85] The displacement of the anterior PAR proteins alleviates the inhibitory effects of PKC-3 and allows PAR-2 loading on the posterior cortex.[86] In the absence of cortical flow, polarity is still established by a redundant microtubule-mediated mechanism of PAR-2 loading on the posterior cortex.[87–89] Microtubules nucleated by the paternal centrosome locally protect PAR-2 from phosphorylation by aPKC, allowing PAR-2 to access the cortex near these microtubules.[89] Recruitment of PAR-1 and LGL-1 by PAR-2 to the cortex and subsequent dissociation of PAR-3 from the cortex by PAR-1 phosphorylation allows the posterior domain to expand.[79,80,86,89] The two established polarity domains are maintained by a positive feedback loop at both sides and by mutual exclusion mechanisms. The feedback loop on the anterior side is dependent on CDC-42, which interacts with PAR-6 and stabilizes anterior PAR-6 localization.[20] At the posterior side PAR-1 created a cortical domain free of PKC-3 allowing localization of more PAR-2 and PAR-1 to this domain. The mutual exclusion mechanisms prevent mixing of the anterior and posterior proteins. Phosphorylation of PAR-3 by PAR-1 excludes the PAR-3/PAR-6/PKC-3 complex from the posterior domain and PAR-2 is excluded from the anterior domain by PKC-3 phosphorylation. Probably the balance between the positive feedback loops and the mutual exclusion mechanisms maintains the polarity at a steady state. The importance of the balance between the different polarity regulators is shown by observations that the absence of one of the regulators can be partially compensated for by altering the protein levels of another regulator. For example, the lethal phenotype of *par-2* null mutants can be partially rescued by decreasing PAR-6 levels or increasing LGL-1 levels.[76,80,81,90]

Epithelial polarity

C. elegans contains a number of different epithelial tissues, such as the epidermis, the intestine, the pharynx, the vulval epithelium, and the spermatheca. As in other organisms, *C. elegans* epithelia are characterized by apical-basal polarity and the presence of an apical junctional complex that prevents paracellular passage of molecules and provides mechanical linkage between the cells. Whereas in mammals distinct tight junctions and adherens

junctions can be observed by EM, the *C. elegans* apical junction (CeAJ) is a single electron-dense structure that consists of two domains: apically the Cadherin Catenin Complex (CCC), consisting of HMR-1 (E-cadherin), HMP-1 (α-catenin), and HMP-2 (β-catenin), and basally the DLG-1 AJM-1 Complex (DAC), consisting of DLG-1, the homolog of *Drosophila* Discs large, and the *C. elegans*-specific protein AJM-1.[91] An overview of the localization of polarity and junctional proteins in *C. elegans* epithelia is shown in Figure 2.

Digestive tract

The two main components of the digestive tract are the pharynx (foregut) and the intestine. The pharynx is the organ through which bacteria are taken up, concentrated, ground, and transported to the intestine. The lumen of the pharynx is lined by epithelial cells, made up of multiple cell types. The bulk of the pharynx is made up by muscle cells. Marginal cells separate the muscle cells from each other. Additional cell types include neurons and gland cells. The pharyngeal epithelium displays apical-basal polarity along the radial axis of the animal, with the apical side facing the lumen. The entire pharynx structure is surrounded by a basal lamina.[92] The intestinal tube is made up of nine rings of cells, the most anterior ring consists of four cells and all other rings of two cells, and the tube twists around its anterior-posterior axis.[93] The complete intestine arises from the E blastomere and only this tissue descends from this cell. After cell division, the two daughter cells of the E blastomeres, the endoderm precursor cells, ingress, a process driven by apical constriction.[94] A series of divisions then leads to the formation of the final 20 intestinal cells. Some of the nuclei go through an extra round of division after the 20 cells are formed to produce a final variable number of up to 34 nuclei when the intestine is fully developed. Apical-basal polarity is established early on during the development of the intestine. At the 1.5-fold stage, foci containing PAR-3 and PAR-6 migrate to the future lateral cortex of the cells.[14] These foci deliver a set of junctional and polarity proteins to the cortex including HMR-1, HMP-1, PAR-6, and PKC-3.[53] In absence of PAR-3, junctional and polarity proteins HMR-1, HMP-1, PAR-6, PKC-3, DLG-1, AJM-1, EAT-20, and LET-413 were mislocalized.[53] A similar process with PAR-3 and PAR-6 containing foci is observed during the development of the pharynx.[95] Little is known about the exact process of polarity establishment in the intestinal cells. It has been shown that most polarity regulators localize in the expected polarized fashion once polarity has been set up. AJM-1 and DLG-1 co-localize to the CeAJ, HMP-1 also localizes to the CeAJ, PKC-3 and CRB-

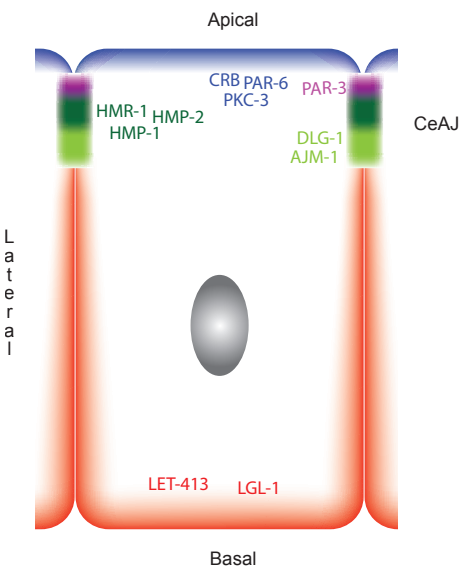


Figure 2. Summarizing overview of the localization of polarity proteins in *C. elegans* epithelial cells.

1 localize to the apical cortex and LET-413 localizes to the basolateral cortex.[50] A similar localization pattern was observed for the pharyngeal cells, where AJM-1, DLG-1, and HMR-1 localize to the CeAJ and PAR-3 and PAR-6 localize to the apical cortex.[92,95,96] Subcellular localization of the polarity proteins is regulated by several mutual dependencies. Depletion of DLG-1 results in mislocalized AJM-1, localization of both DLG-1 and HMP-1 proteins depends on LET-413, and absence of PAR-6 causes mislocalization of PKC-3 and junctional proteins.[28,50]

Seam cells

The *C. elegans* epidermis consists of a single layer of epithelial cells that secrete a collagenous cuticle from its apical, outward facing surface. The first epidermal cells arise on the dorsal side of the embryo and spread out over the embryo to seal it. During embryogenesis epidermal cells fuse to form several syncytia. Expansion of the epidermis during larval stages involves the adding of new nuclei to the syncytia by the stem cell-like seam cells.[97] These seam cells are present in two lateral rows of cells. Most seam cells divide asymmetrically at the beginning of each larval stage and once symmetrically during the 2nd larval stage. The asymmetric division is controlled by Wnt signaling.[98–100] After each round of asymmetric cell division the smaller anterior cell fuses with the largest of the eleven epidermal syncytia, hyp-7, which will contain a final number of 139 nuclei.[97]

In addition to the anterior-posterior Wnt-dependent polarity, the seam cells also possess apical-basal polarity. The Crumbs homolog EAT-20 is located at the apical cortex, while LET-413 localizes to the basolateral cortex of the seam cells.[52,101] AJM-1 is present at the CeAJs between the seam cells and between the seam cells and the syncytia.[9] Loss of PAR-3 or CDC-42 results in loss of cell-cell contacts between the seam cells.[9]

Other epithelia

Cell polarity was studied to some extent in other *C. elegans* epithelia. In the embryonic epidermal cells it was shown that HMR-1 and AJM-1 localize to distinct parts of the CeAJ and that they do not colocalize. HMP-1 and DLG-1 were also observed at the junctions and it was shown that the formation of the DAC was delayed in LET-413 depleted embryos.[102,103] In the vulva epithelium colocalization of AJM-1 and LIN-7 at the CeAJ was shown and apical PAR-6 and basolateral LET-413 localization was observed.[64,104] Another epithelial tissue is formed by the spermathecal precursor cells, which are born during larval development and differentiate into an epithelial tube for the storage of sperm. The apically localized PAR-3, PAR-6, and PKC-3 are all necessary for polarization of the spermatheca. AJM-1, DLG-1, and HMP-1 localize to the junctions, while LET-413 localizes basolaterally.[105,106] Loss of LET-413 leads to mislocalization of AJM-1, PAR-3, and LET-413, while loss of DLG-1 only leads to disruption of AJM-1 localization pattern and does not alter the subcellular localization pattern of PAR-3 and LET-413.[106]

Other polarized tissues

A number of tissues besides epithelia or the one cell embryo are polarized or require one or more cortical polarity regulators for their functioning. An example is the excretory canal, a seamless tube formed by a single cell, which has an apical domain facing the lumen and

an outer basal domain. We observed localization of the novel Crumbs homolog CRB-3 to the canal (Chapter 5). However, the role of cortical polarity regulators in this tissue has not been examined. An example of a cell that relies on the polarity machinery for its functioning is the oocyte. During oocyte development, yolk produced and secreted by the intestine is taken up by the oocytes through an endocytic mechanism that requires the activity of several polarity regulators, including CDC-42, PAR-3, PAR-6, and PKC-3. The same study demonstrated that these polarity regulators are essential for endocytosis by the *C. elegans* coelomocytes, large macrophage-like scavenger cells that are especially active in endocytic uptake of fluid phase molecules.[107] Finally, neuronal Q cell migration and axon guidance is disrupted in animals treated with *par-3* RNAi or *cdc-42* RNAi.[9]

Role of Crumbs and Scribble protein groups in *C. elegans*

While the PAR proteins were identified in *C. elegans* and extensively studied, the roles of the Crumbs and Scribble groups of polarity regulators have received relatively little attention in this organism. Of the Crumbs complex components, only Crumbs itself has been examined. Two Crumbs homologs, CRB-1 and EAT-20 have been described, but inactivation of the corresponding genes, individually or together, does not cause overt defect in epithelial polarity.[50,51] An indication for a more subtle role in cell polarity for CRB-1 comes from studies examining the roles of the *C. elegans* Scribble homolog LET-413 and the *C. elegans* α-catenin homolog HMP-1 in positioning of DLG-1. Depletion of LET-413 results in disrupted positioning of DLG-1, while DLG-1 localization appears normal in *let-413 hmp-1* double knock down embryos. Triple *let-413 hmp-1 crb-1* RNAi leads to a similar phenotype as *let-413* RNAi.[51] Thus, there might be redundancy between the Cadherin-Catenin complex (CCC) and CRB-1. Three candidate Sdt genes exist in *C. elegans*: *magu-1*, *magu-2*, and *magu-3*.[38,63] The roles of these proteins in cell polarity in *C. elegans* have not been investigated. A likely null mutant of MAGU-2, the closest homolog of human PALS1, does not exhibit any obvious phenotype. No clear *C. elegans* homolog of PATJ exists. PATJ consists of an L27 domain followed by multiple PDZ domains (10 in human). The predicted protein MPZ-1 has 10 PDZ domains, while no other proteins have >5 PDZ domains. However, MPZ-1 lacks an L27 domain and appears to act in serotonin signaling pathway, interacting with the serotonin receptor SER-1.[108] The protein also does not appear to show apical localization. Thus, it remains unclear if a *C. elegans* PATJ homolog exists. Finally, as described above, LIN-7 was identified in *C. elegans* as a junctional protein required for the proper basal localization of the LET-23 EGF receptor.[64] However, interactions with other Crumbs components have not been investigated.

C. elegans has single homologs of each of the members of the Scribble group. The *C. elegans* Scribble homolog, LET-413, localizes to the basolateral side of epithelial cells and mediates the apical compaction of junctional proteins, and the apical localization of the intestinal terminal web.[51,101,109] How LET-413 accomplishes this role is unknown, and no protein interaction partners have been identified. The single *C. elegans* DLG homolog, DLG-1, is required for the formation of apical junctions, together with the *C. elegans* specific protein AJM-1.[50,109–111] Finally, the role of the Lgl homolog, LGL-1, has only recently begun to be

elucidated, when it was found that LGL-1 plays a redundant role with PAR-2 in establishment of polarity in the one-cell embryo.[80,81,112]

SCOPE OF THIS THESIS

With the identification of several global regulators of cell polarity, important new questions arise. For example, how does a general polarity establishment machinery bring about such a varied set of polarized cells such as neurons, migrating cells, endotubes, intestinal epithelium, and glands? The polarity regulators described here likely interact with different downstream proteins in different tissues to establish distinct polarity dependent characteristics. To identify the downstream components of the polarity proteins, we developed a tissue-specific affinity purification/mass spectrometry (AP/MS) approach for *C. elegans* (**Chapter 1**).

In addition to identifying the components that mediate polarity establishment in different tissues, it is important to gain a detailed mechanistic understanding of the proteins that regulate polarity. In **Chapter 2**, I describe a systematic yeast two-hybrid based approach for the identification of protein interaction domains. We expand the concept of Y2H-based interaction domain mapping to the human genome-wide level.

To study polarity regulation in a situation as endogenous as possible it is essential to be able to create precise deletions and mutations in the genome. This allows examining a phenotype in absence of (part) of a protein or the phenotype of a mutated version, and adding sequences to or replace sequences of the genome to enable endogenous tagging to follow the dynamics of proteins and to engineer targeted mutations in genes, for example to examine the importance of phosphorylation sites. Precise genome editing was until last year not possible for *C. elegans*. In **Chapter 3**, I describe our development of the CRISPR/Cas9 system for genome engineering in *C. elegans*. We and others simultaneously adapted this system for *C. elegans*, and in **Chapter 4** we provide an overview of the similarities and differences between the approaches taken by each group.

Finally, in **Chapter 5** we examine the role of the Crumbs complex in *C. elegans* development. We identified and characterized a novel third Crumbs homolog, and used our newly developed CRISPR/Cas9 approach to generate a triple knockout lacking all three Crumbs homologs. Surprisingly, animals lacking all three Crumbs homologs are viable. Our results indicate a non-essential role in polarity establishment for the *C. elegans* Crumbs family.

The approaches we developed have already yielded novel insights into polarity establishment in *C. elegans*, and are key enabling technologies to continuing this line of research in the future. In particular the development of CRISPR/Cas9 for precise genome engineering has the potential to be a game changing technology for mechanistic studies of proteins involved in any process.

ACKNOWLEDGEMENTS

I thank Mike Boxem and Sander van den Heuvel for critically reading this manuscript.

REFERENCES

1. Kemphues, K. J., Priess, J. R., Morton, D. G. & Cheng, N. S. Identification of genes required for cytoplasmic localization in early *C. elegans* embryos. Cell 52, 311–20 (1988).
2. Kirby, C., Kusch, M. & Kemphues, K. Mutations in the *par* genes of *Caenorhabditis elegans* affect cytoplasmic reorganization during the first cell cycle. Dev. Biol. 142, 203–15 (1990).
3. Watts, J. L., Morton, D. G., Bestman, J. & Kemphues, K. J. The *C. elegans par-4* gene encodes a putative serine-threonine kinase required for establishing embryonic asymmetry. Development 127, 1467–75 (2000).
4. Morton, D. G. *et al*. The *Caenorhabditis elegans par-5* gene encodes a 14-3-3 protein required for cellular asymmetry in the early embryo. Dev. Biol. 241, 47–58 (2002).
5. Tabuse, Y. *et al*. Atypical protein kinase C cooperates with PAR-3 to establish embryonic polarity in *Caenorhabditis elegans*. Development 125, 3607–14 (1998).
6. Qiu, R. G., Abo, A. & Steven Martin, G. A human homolog of the *C. elegans* polarity determinant Par-6 links Rac and Cdc42 to PKCzeta signaling and cell transformation. Curr. Biol. 10, 697–707 (2000).
7. Kay, A. J. & Hunter, C. P. CDC-42 regulates PAR protein localization and function to control cellular and embryonic polarity in *C. elegans*. Curr. Biol. 11, 474–81 (2001).
8. Gotta, M., Abraham, M. C. & Ahringer, J. CDC-42 controls early cell polarity and spindle orientation in *C. elegans*. Curr. Biol. 11, 482–8 (2001).
9. Welchman, D. P., Mathies, L. D. & Ahringer, J. Similar requirements for CDC-42 and the PAR-3/PAR-6/PKC-3 complex in diverse cell types. Dev. Biol. 305, 347–57 (2007).
10. Goldstein, B. & Macara, I. G. The PAR proteins: fundamental players in animal cell polarization. Dev. Cell 13, 609–22 (2007).
11. St Johnston, D. & Ahringer, J. Cell polarity in eggs and epithelia: parallels and diversity. Cell 141, 757–74 (2010).
12. Motegi, F. & Seydoux, G. The PAR network: redundancy and robustness in a symmetry-breaking system. Phylosophical Trans. R. Soc. Biol. Sci. 368, (2013).
13. McCaffrey, L. M. & Macara, I. G. Epithelial organization, cell polarity and tumorigenesis. Trends Cell Biol. 21, 727–35 (2011).
14. Feldman, J. L. & Priess, J. R. A role for the centrosome and PAR-3 in the hand-off of MTOC function during epithelial polarization. Curr. Biol. 22, 575–82 (2012).
15. Ramahi, J. S. & Solecki, D. J. in Cell. Mol. Control Neuronal Migr. (Nguyen, L. & Hippenmeyer, S.) 800, 113–31 (Springer Netherlands, 2014).
16. Baas, A. F. *et al*. Complete polarization of single intestinal epithelial cells upon activation of LKB1 by STRAD. Cell 116, 457–66 (2004).
17. Joberty, G., Petersen, C., Gao, L. & Macara, I. G. The cell-polarity protein Par6 links Par3 and atypical protein kinase C to Cdc42. Nat. Cell Biol. 2, 531–9 (2000).
18. Lin, D. *et al*. A mammalian PAR-3-PAR-6 complex implicated in Cdc42/Rac1 and aPKC signalling and cell polarity. Nat. Cell Biol. 2, 540–7 (2000).
19. Suzuki, A. *et al*. Atypical protein kinase C is involved in the evolutionarily conserved par protein complex and plays a critical role in establishing epithelia-specific junctional structures. J. Cell Biol. 152, 1183–96 (2001).
20. Aceto, D., Beers, M. & Kemphues, K. J. Interaction of PAR-6 with CDC-42 is required for maintenance but not establishment of PAR asymmetry in *C. elegans*. Dev. Biol. 299, 386–97 (2006).
21. Li, J. *et al*. Binding to PKC-3, but not to PAR-3 or to a conventional PDZ domain ligand, is required for PAR-6 function in *C. elegans*. Dev. Biol. 340, 88–98 (2010).
22. Etemad-Moghadam, B., Guo, S. & Kemphues, K. J. Asymmetrically distributed PAR-3 protein contributes to cell polarity and spindle alignment in early *C. elegans* embryos. Cell 83, 743–52 (1995).
23. Izumi, Y. *et al*. An atypical PKC directly associates and colocalizes at the epithelial tight junction with ASIP, a mammalian homologue of *Caenorhabditis elegans* polarity protein PAR-3. J. Cell Biol. 143, 95–106 (1998).

24. Morais-de-Sá, E., Mirouse, V. & St Johnston, D. aPKC phosphorylation of Bazooka defines the apical/lateral border in *Drosophila* epithelial cells. Cell 141, 509–23 (2010).
25. Harris, T. J. C. & Peifer, M. The positioning and segregation of apical cues during epithelial polarity establishment in *Drosophila*. J. Cell Biol. 170, 813–23 (2005).
26. Walther, R. F. & Pichaud, F. Crumbs/DaPKC-dependent apical exclusion of Bazooka promotes photoreceptor polarity remodeling. Curr. Biol. 20, 1065–74 (2010).
27. Nam, S.-C. & Choi, K.-W. Interaction of Par-6 and Crumbs complexes is essential for photoreceptor morphogenesis in *Drosophila*. Development 130, 4363–72 (2003).
28. Totong, R., Achilleos, A. & Nance, J. PAR-6 is required for junction formation but not apicobasal polarization in *C. elegans* embryonic epithelial cells. Development 134, 1259–68 (2007).
29. Jürgens, G., Wieschaus, E., Nüsslein-Volhard, C. & Kluding, H. Mutations affecting the pattern of the larval cuticle in *Drosophila melanogaster* II. Zygotic loci on the third chromosome. Dev. Biol. 193, 283–95 (1984).
30. Knust, E., Tepass, U. & Wodarz, A. crumbs and stardust, two genes of *Drosophila* required for the development of epithelial cell polarity. Development 261–8 (1993).
31. Bhat, M. A. *et al*. Discs Lost, a novel multi-PDZ domain protein, establishes and maintains epithelial polarity. Cell 96, 833–45 (1999).
32. Bachmann, A., Grawe, F., Johnson, K. & Knust, E. *Drosophila* Lin-7 is a component of the Crumbs complex in epithelia and photoreceptor cells and prevents light-induced retinal degeneration. Eur. J. Cell Biol. 87, 123–36 (2008).
33. Knust, E. *et al*. EGF homologous sequences encoded in the genome of *Drosophila melanogaster*, and their relation to neurogenic genes. EMBO J. 6, 761–6 (1987).
34. Tepass, U., Theres, C. & Knust, E. crumbs encodes an EGF-like protein expressed on apical membranes of *Drosophila* epithelial cells and required for organization of epithelia. Cell 61, 787–99 (1990).
35. Wodarz, A., Hinz, U., Engelbert, M. & Knust, E. Expression of crumbs confers apical character on plasma membrane domains of ectodermal epithelia of *Drosophila*. Cell 82, 67–76 (1995).
36. Izaddoost, S., Nam, S.-C., Bhat, M. A., Bellen, H. J. & Choi, K.-W. *Drosophila* Crumbs is a positional cue in photoreceptor adherens junctions and rhabdomeres. Nature 416, 178–83 (2002).
37. Pellikka, M. *et al*. Crumbs, the *Drosophila* homologue of human CRB1/RP12, is essential for photoreceptor morphogenesis. Nature 416, 143–9 (2002).
38. Assémat, E., Bazellières, E., Pallesi-Pocachard, E., Le Bivic, A. & Massey-Harroche, D. Polarity complex proteins. Biochim. Biophys. Acta 1778, 614–30 (2008).
39. Bachmann, A., Schneider, M., Theilenberg, E., Grawe, F. & Knust, E. *Drosophila* Stardust is a partner of Crumbs in the control of epithelial cell polarity. Nature 414, 638–43 (2001).
40. Hong, Y., Stronach, B., Perrimon, N., Jan, L. Y. & Jan, Y. N. Drosophila Stardust interacts with Crumbs to control polarity of epithelia but not neuroblasts. Nature 414, 634–8 (2001).
41. Roh, M. H. *et al*. The Maguk protein, Pals1, functions as an adapter, linking mammalian homologues of Crumbs and Discs Lost. J. Cell Biol. 157, 161–72 (2002).
42. Bachmann, A. et al. Cell type-specific recruitment of *Drosophila* Lin-7 to distinct MAGUK-based protein complexes defines novel roles for Sdt and Dlg-S97. J. Cell Sci. 117, 1899–909 (2004).
43. Bulgakova, N. A., Kempkens, O. & Knust, E. Multiple domains of Stardust differentially mediate localisation of the Crumbs-Stardust complex during photoreceptor development in *Drosophila*. J. Cell Sci. 121, 2018–26 (2008).
44. Bulgakova, N. A. & Knust, E. The Crumbs complex: from epithelial-cell polarity to retinal degeneration. J. Cell Sci. 122, 2587–96 (2009).
45. Richard, M., Grawe, F. & Knust, E. DPATJ plays a role in retinal morphogenesis and protects against light-dependent degeneration of photoreceptor cells in the *Drosophila* eye. Dev. Dyn. 235, 895–907 (2006).
46. Den Hollander, A. I. *et al*. Mutations in a human homologue of *Drosophila* crumbs cause retinitis pigmentosa (RP12). Nat. Genet. 23, 217–21 (1999).

47. Van den Hurk, J. A. J. M. *et al.* Characterization of the Crumbs homolog 2 (CRB2) gene and analysis of its role in retinitis pigmentosa and Leber congenital amaurosis. Mol. Vis. 11, 263–73 (2005).
48. Makarova, O., Roh, M. H., Liu, C.-J., Laurinec, S. & Margolis, B. Mammalian Crumbs3 is a small transmembrane protein linked to protein associated with Lin-7 (Pals1). Gene 302, 21–9 (2003).
49. Lemmers, C. *et al.* CRB3 binds directly to Par6 and regulates the morphogenesis of the tight junctions in mammalian epithelial cells. 15, 1324–33 (2004).
50. Bossinger, O., Klebes, A., Segbert, C., Theres, C. & Knust, E. Zonula adherens formation in *Caenorhabditis elegans* requires *dlg-1*, the homologue of the *Drosophila* gene discs large. Dev. Biol. 230, 29–42 (2001).
51. Segbert, C., Johnson, K., Theres, C., van Fürden, D. & Bossinger, O. Molecular and functional analysis of apical junction formation in the gut epithelium of *Caenorhabditis elegans*. Dev. Biol. 266, 17–26 (2004).
52. Shibata, Y., Fujii, T., Dent, J. A., Fujisawa, H. & Takagi, S. EAT-20, a novel transmembrane protein with EGF motifs, is required for efficient feeding in *Caenorhabditis elegans*. Genetics 154, 635–46 (2000).
53. Achilleos, A., Wehman, A. M. & Nance, J. PAR-3 mediates the initial clustering and apical localization of junction and polarity proteins during *C . elegans* intestinal epithelial cell polarization. Development 137, 1833–42 (2010).
54. Wang, Q., Chen, X., Margolis, B. & Biotinylation, C. S. PALS1 regulates E-cadherin trafficking in mammalian epithelial cells. Mol. Biol. Cell 18, 874–885 (2007).
55. Cao, X. *et al.* PALS1 specifies the localization of ezrin to the apical membrane of gastric parietal cells. J. Biol. Chem. 280, 13584–92 (2005).
56. Van Rossum, A. G. S. H. *et al.* Pals1/Mpp5 is required for correct localization of Crb1 at the subapical region in polarized Muller glia cells. Hum. Mol. Genet. 15, 2659–72 (2006).
57. Ozçelik, M. *et al.* Pals1 is a major regulator of the epithelial-like polarization and the extension of the myelin sheath in peripheral nerves. J. Neurosci. 30, 4120–31 (2010).
58. Wei, X. & Malicki, J. nagie oko, encoding a MAGUK-family protein, is essential for cellular patterning of the retina. Nat. Genet. 31, 150–7 (2002).
59. Wei, X., Zou, J., Takechi, M., Kawamura, S. & Li, L. Nok plays an essential role in maintaining the integrity of the outer nuclear layer in the zebrafish retina. Exp. Eye Res. 83, 31–44 (2006).
60. Hong, Y., Ackerman, L., Jan, L. Y. & Jan, Y.-N. Distinct roles of Bazooka and Stardust in the specification of *Drosophila* photoreceptor membrane architecture. Proc. Natl. Acad. Sci. U. S. A. 100, 12712–7 (2003).
61. Berger, S., Bulgakova, N. A., Grawe, F., Johnson, K. & Knust, E. Unraveling the genetic complexity of *Drosophila* stardust during photoreceptor morphogenesis and prevention of light-induced degeneration. Genetics 176, 2189–200 (2007).
62. Park, B. *et al.* PALS1 is essential for retinal pigment epithelium structure and neural retina stratification. J. Neurosci. 31, 17230–41 (2011).
63. Knust, E. & Bossinger, O. Composition and formation of intercellular junctions in epithelial cells. Science 298, 1955–9 (2002).
64. Simske, J. S., Kaech, S. M., Harp, S. A. & Kim, S. K. LET-23 receptor localization by the cell junction protein LIN-7 during *C. elegans* vulval induction. Cell 85, 195–204 (1996).
65. Straight, S. W., Pieczynski, J. N., Whiteman, E. L., Liu, C.-J. & Margolis, B. Mammalian lin-7 stabilizes polarity protein complexes. J. Biol. Chem. 281, 37738–47 (2006).
66. Olsen, O. *et al.* Renal defects associated with improper polarization of the CRB and DLG polarity complexes in MALS-3 knockout mice. J. Cell Biol. 179, 151–64 (2007).
67. Srinivasan, K. *et al.* MALS-3 regulates polarity and early neurogenesis in the developing cerebral cortex. Development 135, 1781–90 (2008).
68. Bilder, D. & Perrimon, N. Localization of apical epithelial determinants by the basolateral PDZ protein Scribble. Nature 403, 676–80 (2000).
69. Bilder, D., Li, M. & Perrimon, N. Cooperative regulation of cell polarity and growth by *Drosophila* tumor suppressors. Science (80-.). 289, 113–6 (2000).

70. Thomas, U., Phannavong, B., Müller, B., Garner, C. C. & Gundelfinger, E. D. Functional expression of rat synapse-associated proteins SAP97 and SAP102 in *Drosophila* dlg-1 mutants: effects on tumor suppression and synaptic bouton structure. Mech. Dev. 62, 161–74 (1997).
71. Roberts, S., Delury, C. & Marsh, E. The PDZ protein discs-large (DLG): the "Jekyll and Hyde" of the epithelial polarity proteins. FEBS J. 279, 3549–58 (2012).
72. Zheng, C.-Y., Seabold, G. K., Horak, M. & Petralia, R. S. MAGUKs, synaptic development, and synaptic plasticity. Neuroscientist 17, 493–512 (2011).
73. Humbert, P. O. *et al*. Control of tumourigenesis by the Scribble/Dlg/Lgl polarity module. Oncogene 27, 6888–907 (2008).
74. Betschinger, J., Eisenhaber, F. & Knoblich, J. A. Phosphorylation-induced autoinhibition regulates the cytoskeletal protein Lethal (2) giant larvae. Curr. Biol. 15, 276–82 (2005).
75. Benton, R. & St Johnston, D. *Drosophila* PAR-1 and 14-3-3 inhibit Bazooka/PAR-3 to establish complementary cortical domains in polarized cells. Cell 115, 691–704 (2003).
76. Watts, J. L. *et al*. *par-6*, a gene involved in the establishment of asymmetry in early *C. elegans* embryos, mediates the asymmetric localization of PAR-3. Development 122, 3133–40 (1996).
77. Motegi, F. & Sugimoto, A. Sequential functioning of the ECT-2 RhoGEF, RHO-1 and CDC-42 establishes cell polarity in Caenorhabditis elegans embryos. Nat. Cell Biol. 8, 978–85 (2006).
78. Guo, S. & Kemphues, K. J. *par-1*, a gene required for establishing polarity in *C. elegans* embryos, encodes a putative Ser/Thr kinase that is asymmetrically distributed. Cell 81, 611–20 (1995).
79. Boyd, L., Guo, S., Levitan, D., Stinchcomb, D. T. & Kemphues, K. J. PAR-2 is asymmetrically distributed and promotes association of P granules and PAR-1 with the cortex in *C. elegans* embryos. Development 122, 3075–84 (1996).
80. Hoege, C. *et al*. LGL can partition the cortex of one-cell *Caenorhabditis elegans* embryos into two domains. Curr. Biol. 20, 1296–303 (2010).
81. Beatty, A., Morton, D. & Kemphues, K. The *C. elegans* homolog of *Drosophila* Lethal giant larvae functions redundantly with PAR-2 to maintain polarity in the early embryo. Development 137, 3995–4004 (2010).
82. Hung, T. J. & Kemphues, K. J. PAR-6 is a conserved PDZ domain-containing protein that colocalizes with PAR-3 in *Caenorhabditis elegans* embryos. Development 126, 127–35 (1999).
83. Cuenca, A. A., Schetter, A., Aceto, D., Kemphues, K. & Seydoux, G. Polarization of the *C. elegans* zygote proceeds via distinct establishment and maintenance phases. Development 130, 1255–65 (2003).
84. Goldstein, B. & Hird, S. N. Specification of the anteroposterior axis in *Caenorhabditis elegans*. Development 122, 1467–74 (1996).
85. Munro, E., Nance, J. & Priess, J. R. Cortical flows powered by asymmetrical contraction transport PAR proteins to establish and maintain anterior-posterior polarity in the early *C. elegans* embryo. Dev. Cell 7, 413–24 (2004).
86. Hao, Y., Boyd, L. & Seydoux, G. Stabilization of cell polarity by the *C. elegans* RING protein PAR-2. Dev. Cell 10, 199–208 (2006).
87. Shelton, C. A., Carter, J. C., Ellis, G. C. & Bowerman, B. The nonmuscle myosin regulatory light chain gene *mlc-4* is required for cytokinesis, anterior-posterior polarity, and body morphology during *Caenorhabditis elegans* embryogenesis. J. Cell Biol. 146, 439–451 (1999).
88. Zonies, S., Motegi, F., Hao, Y. & Seydoux, G. Symmetry breaking and polarization of the *C. elegans* zygote by the polarity protein PAR-2. Development 137, 1669–77 (2010).
89. Motegi, F. *et al*. Microtubules induce self-organization of polarized PAR domains in *Caenorhabditis elegans* zygotes. Nat. Cell Biol. 13, 1–9 (2011).
90. Labbé, J.-C., Pacquelet, A., Marty, T. & Gotta, M. A genomewide screen for suppressors of *par-2* uncovers potential regulators of PAR protein-dependent cell polarity in *Caenorhabditis elegans*. Genetics 174, 285–95 (2006).
91. Costa, M. *et al*. A putative catenin–cadherin system mediates morphogenesis of the *Caenorhabditis elegans* embryo. J. Cell Biol. 141, 297–308 (1998).

92. Portereiko, M. F. & Mango, S. E. Early morphogenesis of the *Caenorhabditis elegans* pharynx. Dev. Biol. 233, 482–94 (2001).
93. Leung, B., Hermann, G. J. & Priess, J. R. Organogenesis of the *Caenorhabditis elegans* intestine. Dev. Biol. 216, 114–34 (1999).
94. Rohrschneider, M. R. & Nance, J. Polarity and cell fate specification in the control of *Caenorhabditis elegans* gastrulation. Genes Dev. 789–796 (2009). doi:10.1002/dvdy.21893
95. Rasmussen, J. P., Somashekar Reddy, S. & Priess, J. R. Laminin is required to orient epithelial polarity in the *C. elegans* pharynx. Development 2060, 2050–60 (2012).
96. Rasmussen, J. P., Feldman, J. L., Reddy, S. S. & Priess, J. R. Cell interactions and patterned intercalations shape and link epithelial tubes in *C. elegans*. PLoS Genet. 9, e1003772 (2013).
97. Chisholm, A. D. & Hsiao, T. I. The *C. elegans* epidermis as a model skin. I: development, patterning, and growth. Wiley Interdiscip. Rev. Dev. Biol. 2, 1–25 (2013).
98. Mizumoto, K. & Sawa, H. Two betas or not two betas: regulation of asymmetric division by beta-catenin. Trends Cell Biol. 17, 465–73 (2007).
99. Goldstein, B., Takeshita, H., Mizumoto, K. & Sawa, H. Wnt signals can function as positional cues in establishing cell polarity. Dev. Cell 10, 391–6 (2006).
100. Yamamoto, Y., Takeshita, H. & Sawa, H. Multiple Wnts redundantly control polarity orientation in *Caenorhabditis elegans* epithelial stem cells. PLoS Genet. 7, e1002308 (2011).
101. Legouis, R. *et al*. LET-413 is a basolateral protein required for the assembly of adherens junctions in *Caenorhabditis elegans*. Nat. Cell Biol. 2, 415–22 (2000).
102. Simske, J. S. *et al*. The cell junction protein VAB-9 regulates adhesion and epidermal morphology in *C. elegans*. Nat. Cell Biol. 5, 619–25 (2003).
103. Pilipiuk, J., Lefebvre, C., Wiesenfahrt, T., Legouis, R. & Bossinger, O. Increased IP 3 / Ca 2 + signaling compensates depletion of LET-413 / DLG-1 in *C . elegans* epithelial junction assembly. Dev. Biol. 327, 34–47 (2009).
104. Hurd, D. D. & Kemphues, K. J. PAR-1 is required for morphogenesis of the *Caenorhabditis elegans* vulva. Dev. Biol. 253, 54–65 (2003).
105. Aono, S., Legouis, R., Hoose, W. A. & Kemphues, K. J. PAR-3 is required for epithelial cell polarity in the distal spermatheca of *C. elegans*. Development 2865–74 (2004). doi:10.1242/dev.01146
106. Pilipiuk, J., Lefebvre, C., Wiesenfahrt, T., Legouis, R. & Bossinger, O. Increased IP3/Ca2+ signaling compensates depletion of LET-413/DLG-1 in *C. elegans* epithelial junction assembly. Dev. Biol. 327, 34–47 (2009).
107. Balklava, Z., Pant, S., Fares, H. & Grant, B. D. Genome-wide analysis identifies a general requirement for polarity proteins in endocytic traffic. Nat. Cell Biol. 9, 1066–73 (2007).
108. Xiao, H. *et al*. SER-1, a *Caenorhabditis elegans* 5-HT2-like receptor, and a multi-PDZ domain containing protein (MPZ-1) interact in vulval muscle to facilitate serotonin-stimulated egg-laying. Dev. Biol. 298, 379–91 (2006).
109. McMahon, L., Legouis, R., Vonesch, J. L. & Labouesse, M. Assembly of *C. elegans* apical junctions involves positioning and compaction by LET-413 and protein aggregation by the MAGUK protein DLG-1. J. Cell Sci. 114, 2265–77 (2001).
110. Köppen, M. *et al*. Cooperative regulation of AJM-1 controls junctional integrity in *Caenorhabditis elegans* epithelia. Nat. Cell Biol. 3, 983–91 (2001).
111. Firestein, B. L. & Rongo, C. DLG-1 is a MAGUK similar to SAP97 and is required for adherens junction formation. Mol. Biol. Cell 12, 3465–75 (2001).
112. Beatty, A., Morton, D. & Kemphues, K. The *C. elegans* homolog of *Drosophila* Lethal giant larvae functions redundantly with PAR-2 to maintain polarity in the early embryo. Development 137, 3995–4004 (2010).

Chapter 1

Identification of tissue-specific protein complexes in *Caenorhabditis elegans*

Selma Waaijers[1], Soenita S. Goerdayal[2,3], Sander van den Heuvel[1], Baris Tursun[4], Javier Muñoz[2,3,5], Albert J. Heck[2,3] & Mike Boxem[1]

[1] Utrecht University, Department of Biology, Padualaan 8, 3584 CH Utrecht, The Netherlands.

[2] Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands.

[3] Netherlands Proteomics Centre, Padualaan 8, 3584 CH Utrecht, The Netherlands.

[4] Berlin Institute for Medical Systems Biology (BIMSB) at Max-Delbrück-Centre for Molecular Medicine (MDC), Robert-Rössle-Strasse 10, Berlin 13125, Germany.

[5] Present address: Proteomics Unit, Spanish National Cancer Research Centre (CNIO) 28029 Madrid, Spain. ProteoRed-ISCIII

ABSTRACT

Affinity purification followed by mass spectrometry (AP/MS) approaches to identify protein complexes have mainly focused on single cell systems. Applying AP/MS approaches in multicellular organisms adds the challenge of revealing the composition of protein complexes that may differ depending on the tissue. We developed a tissue-specific protein purification approach for *Caenorhabditis elegans* based on the *in vivo* biotinylation of Avi-tagged proteins of interest by the bacterial biotin ligase BirA. Purification of biotinylated bait proteins with streptavidin-coated beads allows the identification of interacting proteins by mass spectrometry. Bait proteins are expressed from their native regulatory sequences, while tissue-specific biotinylation is accomplished by expressing BirA from tissue-specific promoters. We developed N- and C-terminal GFP::Avi tags that enable us to follow the localization of the bait proteins *in vivo*. Control experiments expressing GFP::Avi and BirA in two separate tissues confirm the tissue-specific nature of the biotinylation and purification. We further tested our approach by tagging several cortical polarity regulators. GFP::Avi-tagged PAR-3, DLG-1, LET-413, LGL-1, and CDC-42 all displayed the expected subcellular localization patterns, while complementation assays further demonstrate that addition of the GFP::Avi tags does not disrupt protein function. Initial purifications of DLG-1 from intestine and CDC-42 from the seam cell epithelium identified a number of known interaction partners, demonstrating that our approach can identify valid interactions from specific cells or tissues. We are currently investigating the complex composition of these five polarity regulators in two epithelial tissues, as well as the interactions of DLG-1 in neurons.

INTRODUCTION

Knowledge of protein complex composition has had a major impact on our understanding of cellular processes and signal transduction pathways.[1] Insight in the composition of a protein complex can be obtained through affinity purification followed by mass spectrometry (AP/MS). Systematic AP/MS efforts have mostly focused on single cell systems, such as yeast, bacteria, and cultured *Drosophila melanogaster* cells.[2–5] In multicellular organisms, AP/MS approaches face the added complexity that the composition and function of a protein complex involving a particular protein of interest may differ per tissue. Purifying a bait protein from whole animal lysates may not reveal the composition of a normal functional complex but result in identification of members of multiple different complexes. Distinguishing these complexes to better interpret the biological meaning of the identified interactors demands the development of new approaches to purify protein complexes from specific tissues. The nematode *Caenorhabditis elegans* is a widely used multicellular model animal that contains a large variety of differentiated cell types and tissues, including several epithelial tissues, neurons, and muscle. A number of studies have used AP/MS approaches to purify *C. elegans* proteins and identify interaction partners.[6–15] However, affinity purification is not widely used for *C. elegans* and a method for purification of proteins from specific tissues is lacking.

In multicellular organisms, cell polarity is crucial during development to generate different cell types through asymmetric cell division, for guiding cell migration and axonal outgrowth, and for the proper functioning of all epithelia. Epithelia are the most common polarized tissues, which depend on functional specialization of opposing apical and basal surfaces to perform their role as selectively permeable barriers between different body compartments and the environment. Three evolutionary conserved groups of proteins are essential for the regulation of apical-basal polarity in epithelia: the Scribble group (SCRIB/DLG/LGL), the PAR proteins (PAR-3/PAR-6/aPKC) and the Crumbs complex (CRB/PALS1/LIN-7/PATJ).[16] The members of the PAR complex were identified in a screen for defects in asymmetric partitioning of cell fate determinants in the *C. elegans* embryo, and were shown to be important for establishing and maintaining cell polarity in a wide range of cell types in many organisms.[17,18] In the *C. elegans* embryo, PAR-3, PAR-6 and aPKC form a complex that localizes to the anterior pole. In polarized epithelial cells, PAR-6 and aPKC reside in a complex at the apical domain, while PAR-3 is frequently excluded from this domain and found just apical of the apical junctional complex (AJC).[19–22] The Scribble and Crumbs groups, both identified originally in *Drosophila*, play more restricted roles in polarity establishment and are largely specific to the regulation of epithelial polarity. Crumbs complex proteins localize apically, and are often enriched at a region just apical of the AJC.[23] Scribble proteins localize at the lateral or basolateral membranes.[24,25]

A major mechanism through which these three groups of polarity regulators establish and maintain distinct membrane domains is mutual exclusion, where proteins in one domain prevent the invasion of proteins from the other domain.[26,27] However, the mechanisms used to establish polarity appears to vary markedly in different cell types or situations. For example, in *Drosophila*, not all epithelia in which Crb is expressed require Crb to maintain

epithelial polarity.[28] In the *C. elegans* embryo, PAR-3 shows distinct requirements in different epithelial tissues. In intestinal cells, PAR-3 is required for the assembly of cell junctions. In epidermal epithelia however, apical junctions still form in the absence of PAR-3, and PAR-6 promotes junction maturation independently from PAR-3.[29] It is likely that the distinct requirements are reflected by the formation of different protein complexes in various tissues. Though the Crumbs complex has been identified in *Drosophila* as a regulator of cell polarity and is also extensively studied in vertebrates, its role in cell polarity regulation in *C. elegans* is not yet elucidated.[23,30–32] The members of the Scribble group all have homologues in *C. elegans* functioning in cell polarity, but little is known about how LGL-1 and the Scribble homolog, LET-413, carry out their functions.[31,33,34] Most of these polarity regulators possess a high number of protein-protein interaction domains and may act as scaffolds, recruiting downstream partners that are needed to bring about the final epithelial character of the cell.

In order to better understand the regulation of epithelial polarity in different tissues and to identify novel components that act downstream of cortical polarity regulators, we sought to develop a method for the tissue-specific purification of protein complexes from *C. elegans*. We adapted an *in vivo* biotinylation based approach in which an Avi-tagged bait protein is biotinylated in a specific tissue by expression of the biotin ligase BirA from a tissue-specific promoter.[35,36] Biotinylated bait proteins and associated proteins are then precipitated with streptavidin-coated beads, and their identities are determined by mass spectrometry. We demonstrate the applicability of this approach by showing tissue-specific biotinylation of Avi-tagged GFP and identifying a number of previously known protein interactors of DLG-1 and CDC-42 in specific tissues.

RESULTS

A biotinylation-based tissue-specific protein purification approach

To purify proteins from specific *C. elegans* tissues, we adapted a system based on *in vivo* biotinylation of a protein of interest. In this approach, the protein of interest is tagged with the 15 amino acid Avi tag, which can be biotinylated *in vivo* by the BirA biotin ligase from *Escherichia coli*.[35,36] Biotinylated proteins are then purified using streptavidin-coated beads. Importantly, biotinylation only takes place at the intersection of the expression profiles of BirA and the Avi-tagged protein. This allows us to express the protein of interest from its native regulatory sequences, while biotinylation in a specific tissue is accomplished by expressing BirA from tissue-specific promoters. Expression of the bait protein from its native regulatory sequences has three main advantages. First, rescue of mutants can be used to test whether the tagged protein remains functional, and strains can be created that do not express the untagged endogenous protein. Second, only a single transgenic line needs to be created, which can then be crossed to multiple BirA driver lines to purify the protein of interest from different tissues (Fig. 1A). Compared to alternative approaches expressing tagged proteins from tissue-specific promoters, our strategy reduces the number of transgenic strains that need to be laboriously generated. Third, it avoids the risk of expressing the protein in a tissue where it would normally not be expressed.

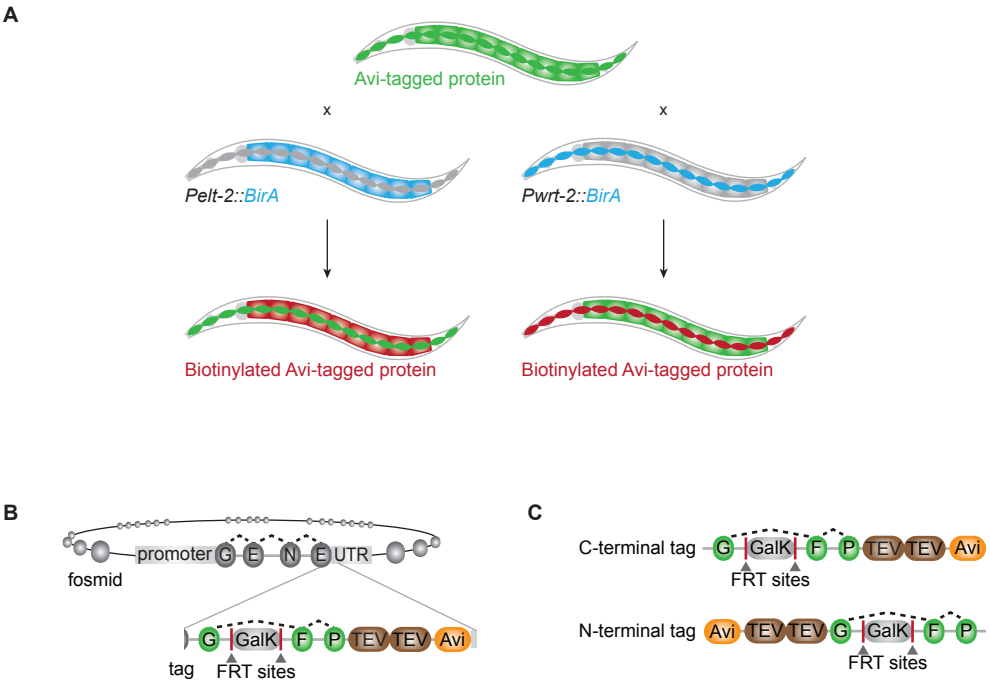


Figure 1. Principles of the approach. (A) By crossing a transgenic *C. elegans* strain expressing the tagged gene of interest (green) with different BirA driver lines (blue), biotinylation of the tagged protein (red) is accomplished in distinct tissues. (B) Recombineering is used to incorporate a tag into a fosmid containing the gene of interest. (C) The C-terminal and N-terminal tag designed for this approach containing GFP, two TEV cleavage sites, and the Avi-tag sequence. *GalK* is a selection marker used during the recombineering. This selection marker is removed by Flip recombination before injection of the construct into animals, leaving a single FRT site in an intron of GFP.

We inserted the tag-encoding sequences into fosmids carrying the gene of interest through recombineering, a homologous recombination-based genetic engineering technique in bacteria.[37] This approach enabled us to integrate our tag into a large region of approximately 30 to 40 kb of genomic DNA that likely contains all the native regulatory sequences of the gene of interest, including the promoter, 3' UTR, and introns (Fig. 1B). We designed N- and C-terminal tags consisting of GFP and the Avi sequence, separated by two Tobacco Etch Virus (TEV) cleavage sites (Fig. 1C, from here on referred to as the GFP::Avi tag). The presence of GFP enables us to examine the expression pattern and subcellular localization of the protein of interest, either to confirm that the tagged protein localizes as expected or to study the localization of proteins whose localization pattern has not yet been (fully) described. A combined GFP and affinity purification tag has been used successfully in *C. elegans*.[7] The addition of TEV cleavage sites was necessary to eliminate biotinylated proteins naturally present in *C. elegans*, which will also be purified by the streptavidin-coated beads. TEV cleavage releases the bait protein and any associated proteins from the beads, while the background of naturally biotinylated proteins stays bound. GFP, TEV, and Avi are separated

by short flexible linkers of five small amino acids, while GFP is separated from the bait protein by a longer flexible linker of 13 small amino acids.

An overview of the entire tagging and AP/MS procedure is shown in Figure 2. GFP::Avi encoding sequences are added to a fosmid containing a gene of interest by recombineering. Next, transgenic strains are generated by germline injection of the resulting fosmids together with reporter plasmids, and subsequently strains with an integrated transgene array are generated by gamma irradiation. Transgenic lines expressing the Avi-tagged bait protein are then crossed to a BirA expressing strain. Whenever possible, transgenic lines were also crossed to a mutant lacking the endogenous protein of interest. This ensures that the tagged protein is fully functional and eliminates the presence of wild-type untagged protein, which would otherwise reduce the fraction of complexes incorporating a tagged protein. The lines are grown at large scale in liquid culture before harvesting, lysis, and purification of biotinylated proteins with streptavidin-coated beads. The bait protein and any bound proteins are then cleaved off the beads using TEV protease, and analyzed by tandem mass spectrometry to determine their identities.

***In vivo* biotinylation and purification are highly tissue-specific**

Tissue-specific expression of BirA should result in purification of the bait protein specifically from that tissue. However, at least two potential problems might give rise to biotinylation – and thus purification – of proteins from different tissues. First, if the expression of BirA is not tightly limited to the tissue of interest, BirA might be expressed at low levels in other tissues. However, many promoters with well documented and highly tissue-specific expression patterns have been identified in *C. elegans*. Second, BirA might biotinylate Avi-tagged proteins after the lysis procedure, when proteins from the entire animal are mixed together. The risk of this occurrence is minimal, as BirA activity requires the presence of bivalent ions, which are chelated by EDTA in the lysis buffer.[38] Nevertheless, we first wished to demonstrate the specificity of biotinylation and purification from a specific tissue.

To test the tissue-specificity of our approach, we generated transgenic *C. elegans* strains expressing cytoplasmic Avi-tagged GFP and BirA, either in the same tissue or in two separate tissues. We generated two strains expressing GFP::Avi from the intestinal *elt-2* promoter and the seam cell-specific *wrt-2* promoter. Both GFP::Avi strains showed high levels of cytoplasmic GFP expression in the expected cells (Fig. 3A). We also generated BirA driver lines expressing *C. elegans* codon optimized and Myc-tagged BirA from the *elt-2* and *wrt-2* promoters. Immunostaining using a Myc-tag specific antibody confirmed expression of BirA in the intestine and seam cells, respectively (Fig. 3A). In addition to these epithelial BirA lines, we generated lines expressing BirA ubiquitously (using the *rps-27* promoter), and in neuronal cells (using the *rgef-1* promoter).

Each of the GFP::Avi strains was crossed with the two epithelial BirA driver strains. Of the four resulting strains, two express the biotin ligase and Avi-tagged GFP in the same tissue (intestine or seam cells), while the other two express BirA and GFP::Avi in two different tissues. To detect the presence of biotinylated GFP::Avi, animals of all four strains were lysed, biotinylated proteins were purified using streptavidin-coated beads, and the presence

of GFP amongst the biotinylated proteins was examined by western blot. All samples showed the expected biotinylation pattern: biotinylation of the Avi-tag when BirA and GFP::Avi are present in the same tissue, and no biotinylation when BirA and GFP::Avi are expressed in separate tissues (Fig. 3B). Thus, *in vivo* biotinylation and protein purification are highly tissue-specific.

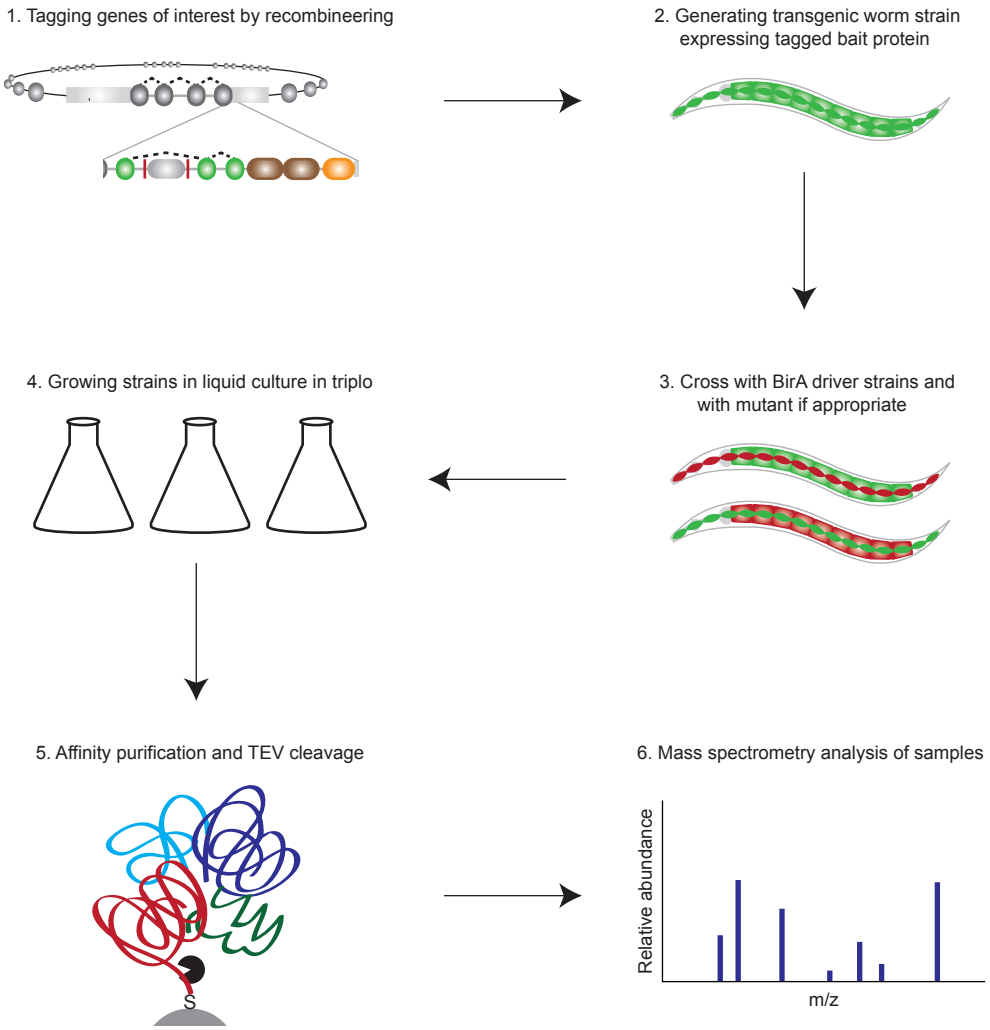


Figure 2. Schematic of workflow. (1) The GFP::Avi tag is added to a gene of interest using recombineering. (2) Transgenic *C. elegans* strains expressing the GFP::Avi-tagged protein are generated by injection followed by gamma irradiation-mediated integration of the extrachromosomal array. (3) Transgenic lines are crossed with strains expressing BirA from a tissue-specific promoter, and with a genetic null mutant if appropriate. (4) The transgenic strains are grown in triplo in liquid culture. (5) Affinity purification is performed on whole-animal lysates. The bait protein with any interacting proteins is subsequently cleaved off the beads by TEV protease. (6) The samples are analyzed by tandem mass spectrometry to identify the proteins it contains.

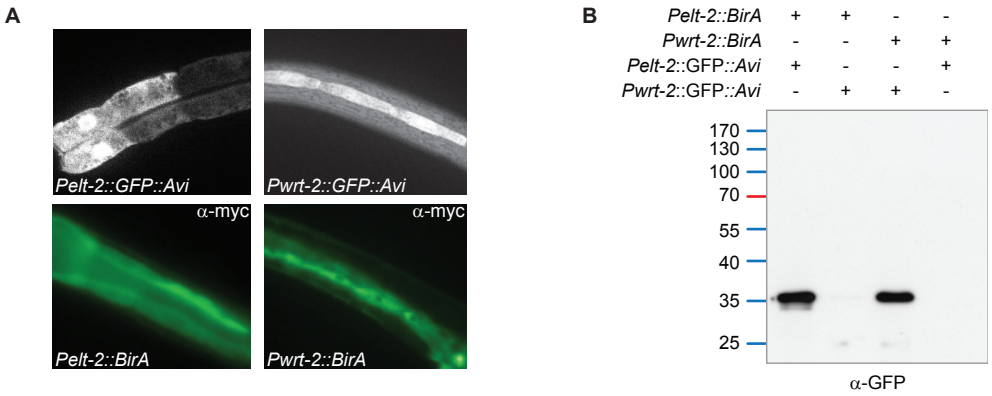


Figure 3. Tissue-specific expression of GFP::Avi and BirA, and *in vivo* biotinylation in distinct tissues. (A) Upper panels: expression pattern of GFP::Avi from the intestinal *elt-2* promoter or the seam cell-specific *wrt-2* promoter, respectively. Bottom panels: immunostaining of strains expressing myc-tagged BirA in the intestine (*Pelt-2*) or the seam cells (*Pwrt-2*), respectively. (B) Western blot probed against GFP on streptavidin pull down samples on strains expressing GFP-2xTEV-Avi and BirA either in the same tissue or in distinct tissues.

Purification of ubiquitously biotinylated DLG-1 identifies known binding partners

We wanted to determine whether affinity purification of a biotinylated bait protein can identify relevant interaction partners. As a test case, we generated a line expressing GFP::Avi tagged DLG-1, the *C. elegans* Discs large homolog, under control of its own regulatory sequences as well as BirA expressed from the ubiquitous *sur-5* promoter. The line was generated by co-injection of the engineered *dlg-1* fosmid and the *Psur-5::BirA* expression construct, followed by gamma irradiation-mediated integration of the extrachromosomal array. The transgenic line, as well as control N2 animals, was grown in large-scale liquid culture. After growth in liquid culture, animals were harvested when most larvae were at the third or fourth larval stage. Both strains were grown in triplicate, to be able to perform three independent AP/MS experiments. The use of multiple independent replicates significantly enhances the possibility of separating *bona fide* protein-protein interactions from nonspecific interactions following mass spectrometry analysis, by taking into account the number of samples in which an interacting protein is present and the number of control samples in which it is absent.

The harvested cultures were lysed by sonication, and biotinylated proteins were purified from the lysates using streptavidin-coated beads. Bait proteins and binding partners were released from the streptavidin beads through TEV protease cleavage. To determine the efficiency of the purification and subsequent release of DLG-1::GFP from the beads by TEV cleavage, we analyzed the sample before and after TEV cleavage for the presence of DLG-1::GFP and biotinylated proteins (Fig. 4). DLG-1::GFP::Avi is biotinylated by BirA (Fig. 4A) and is together with a number of endogenously biotinylated proteins captured by the streptavidin beads. The TEV cleavage releases DLG-1::GFP from the beads, while the biotinylated Avi-tag and endogenous biotinylated proteins remain bound to the beads, and are thus eliminated from mass spectrometry analysis (Fig. 4B).

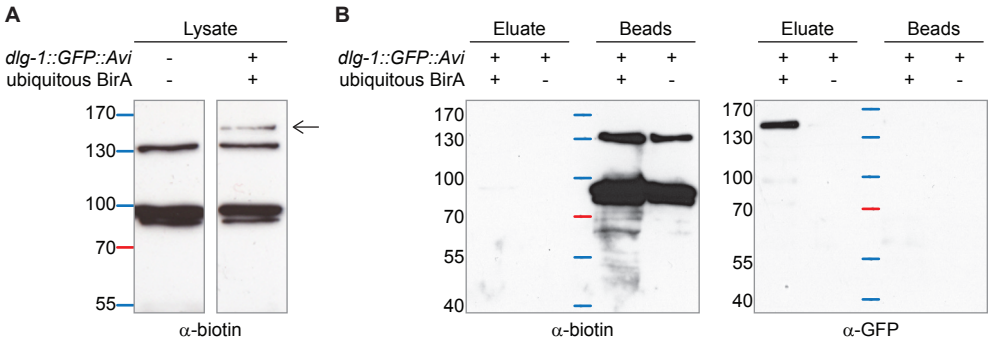


Figure 4. Efficient TEV cleavage of tagged DLG-1. (A) A western blot probed with α-biotin. The background of naturally biotinylated proteins is visible in the first lane containing wild type *C. elegans* lysate. An additional band corresponding to biotinylated DLG-1::GFP::Avi, indicated by the arrow, is visible in the second lane containing lysate of a strain expressing tagged DLG-1 and ubiquitous BirA. (B) Two western blots containing the same samples, one probed with α-biotin, the other with α-GFP. The background of naturally biotinylated proteins remains bound to the streptavidin beads and is visible on the α-biotin western blot in the beads remainder, while tagged DLG-1 is cleaved of and visible on the α-GFP western blot in the eluate.

We next analyzed the samples by tandem mass spectrometry. In each of the DLG-1 purifications, the highest number of identified *C. elegans* peptides matched the bait protein (Table 1). Occasionally we also identified DLG-1 in control samples, though always at very low levels, and we speculate that this is due to retention of a small amount of the highly abundant DLG-1 protein in the HPLC column. We also identified two known DLG-1 interacting proteins: AJM-1 and ATAD-3 (ATAD-3 by personal communication C. Berends and S. van den Heuvel).[39] ATAD-3 was identified in 2 out of 3 samples with an average peptide count of 6 and was not identified in the control samples (Table 1). AJM-1 was identified in 1 out of 3 samples with 1 peptide and not in the control samples (Table 1). Thus, two known interacting proteins were identified specifically in the DLG-1 samples absent from control purifications. While the number of peptides was low, these pilot experiments demonstrate that *in vivo* biotinylation can be used for the identification of relevant protein interaction partners.

GFP::Avi-tagged polarity regulators are functional and show polarized localization

We chose five cell polarity regulators with distinct subcellular localization as bait proteins to further validate our method: the (sub)apical protein PAR-3, the junctional protein DLG-1, the basolateral proteins LET-413 and LGL-1, and the overall cortical protein CDC-42. These

Table 1. Mass spectrometry results of purified DLG-1 when expressed from a ubiquitous promoter and of wild type controls in triplo.

Protein	# PSMs control 1	# PSMs control 2	# PSMs control 3	# PSMs sample 1	# PSMs sample 2	# PSMs sample 3	In x/x controls	In x/x samples
DLG-1::GFP::Avi	17	-	26	55	477	564	2/3	3/3
ATAD-3	-	-	-	-	10	2	0/3	2/3
AJM-1	-	-	-	-	-	1	0/3	1/3

five proteins all contain multiple protein-protein interaction domains and for some of these proteins, interacting partners have already been identified in *C. elegans*. In addition to interaction of DLG-1 with ATAD-3 and the *C. elegans*-specific protein AJM-1 described above, PAR-3 can interact with PKC-3 and PAR-6, and CDC-42 binds to PAR-6.[39–41] For the Scribble homolog LET-413 and LGL-1, interacting proteins have only been found in other organisms. For four of these proteins, PAR-3, DLG-1, LET-413, and LGL-1, we added the GPF::Avi tag to the C-terminus, as C-terminal tags are in general less disruptive to protein function than N-terminal tags.[42] As it was already shown that *C. elegans* GFP::CDC-42 localizes to its expected cortical location, we used the N-terminal tag for CDC-42.[41] For all five genes, we successfully engineered fosmids to incorporate the GFP::Avi tag and generated integrated transgenic lines by injection and gamma irradiation. We examined the expression pattern of the tagged bait proteins in the intestine and seam cells and found that all proteins had the expected subcellular localization (Fig. 5). In both tissues we observed junctional localization of DLG-1::GFP::Avi, basolateral localization of LET-413::GFP::Avi and LGL-1::GFP::Avi, and cortical localization of Avi::GFP::CDC-42. PAR-3::GFP-2xTEV-Avi was visible at the apical surface of the seam cells. However, no tagged PAR-3 was detected in the intestinal epithelium. In addition to the epithelial localization, we also observed expression of DLG-1::GFP::Avi in the ventral nerve cord (Fig. 5C), which is consistent with the expression in neurons of the human homolog PSD95.[43]

To further test the functionality of the fusion proteins, we crossed each of the transgenic lines with lines carrying a mutant allele of the corresponding endogenous gene, if available. Rescue of the mutant phenotype is the best indicator of functionality of a transgene. In addition, transgene expression in a null mutant background allows maximum incorporation of tagged protein in protein complexes, as there is no competition of wild-type untagged protein. To test the rescuing capability of DLG-1::GFP::Avi, we crossed the transgenic line with a line carrying the *dlg-1(ok318)* deletion allele, which results in embryonic arrest at the 2-fold stage. Homozygous *dlg-1(ok318)* mutants carrying the *dlg-1::GFP::Avi* fosmid were fully viable, demonstrating that the DLG-1::GFP::Avi protein is functional. All further experiments with DLG-1::GFP::Avi were carried out in the *dlg-1(ok318)* background. For *lgl-1*, it was not possible to examine the rescuing capacity of the *lgl-1::GFP::Avi* fosmid, as loss of *lgl-1* alone does not affect cell polarity or viability.[34] Nevertheless, we crossed our *lgl-1::GFP::Avi* transgenic line with the *lgl-1(tm2616)* deletion allele, which causes a premature stop and presumably prevents incorporation of untagged LGL-1 protein into complexes. For *let-413*, no molecular null allele is available. We therefore crossed our *let-413::GFP::Avi* transgenic line with a strain carrying the *let-413(s128)* missense mutation, which causes embryonic lethality. Expression of LET-413::GFP::Avi fully rescued this phenotype. However, the *let-413(s128)* allele was linked in cis to *dpy-11* and *unc-42* mutations, which caused the strain to grow slowly. We therefore decided not to use a mutant background for the LET-413 purifications. *cdc-42* and *par-3* null alleles are early embryonic lethal.[18,44] Transgenes are usually silenced in the germline and are therefore unlikely to rescue these phenotypes. Consequently, we did not use mutant backgrounds for these two bait proteins.

In summary, each of the 5 tagged proteins exhibit the expected subcellular localization pattern, and the two transgenes that we could test in complementation assays fully rescue the phenotype of the corresponding mutant alleles. Thus, the experimental design appears well suited for the tagging of proteins of interest, without interfering with their function.

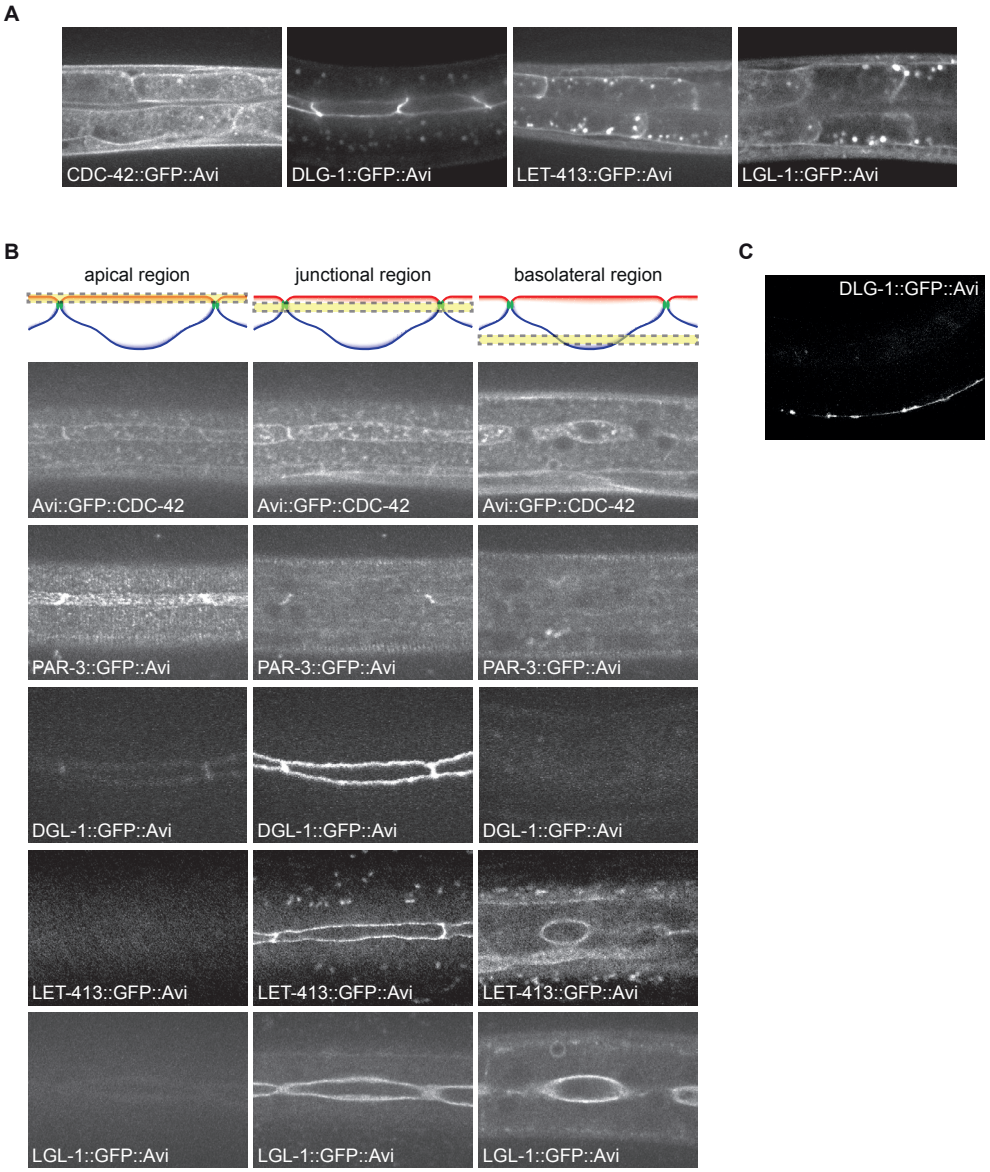


Figure 5. Spinning disc confocal immunofluorescence microscopy images, showing localization of GFP::Avi tagged proteins in intestinal cells (A) and seam cells (B), as well as DLG-1::GFP::Avi localization in the nerve cord (C). For the protein localization in the seam cells, three different regions of the cells are depicted: the apical region, the region where the junctions are positioned, and the basolateral region.

Tissue-specific purification of DLG-1 and CDC-42 identifies known binding partners

Each of the five transgenic *C. elegans* strains expressing a tagged polarity protein was crossed with three different BirA driver strains, which express BirA in the intestine from the *elt-2* promoter, in seam cells through the *wrt-2* promoter, and ubiquitously through the *rps-27* promoter, respectively. As described above, the strains expressing DLG-1::GFP::Avi and LGL-1::GFP::Avi also carried a putative null mutation in the corresponding endogenous gene. The DLG-1::GFP::Avi expressing strain was also combined with neuronal BirA expressed from the *rgef-1* promoter, because of the observed neuronal expression of DLG-1. We also generated a transgenic strain expressing cytoplasmic GFP::Avi from the ubiquitous *rps-27* promoter, which was crossed with each of the four BirA driver strains. These latter strains serve as control lines to identify common contaminant proteins as well as proteins that bind to the GFP::Avi tag. In total 20 transgenic strains were generated for purification: the five tagged bait proteins and the control strain, each combined with BirA expressed from three different promoters, and the DLG-1 and GFP::Avi control strain combined with neuronal BirA.

The 20 strains were grown in triplicate in large-scale liquid culture and harvested at the L3/L4 stage as above. All of the harvested cultures were lysed by sonication, and biotinylated proteins were purified from the lysates using streptavidin-coated beads. Next, bait proteins and binding partners were released from the streptavidin beads through TEV protease cleavage.

To test the ability of our approach to identify *bona fide* protein interactions from specific tissues, we initially focused our attention on DLG-1 and CDC-42. We analyzed three replicates of CDC-42 purified from the seam cells, as well as a single sample of DLG-1 purified from the intestine by tandem mass spectrometry. As a negative control, we analyzed three replicates of cytoplasmic GFP::Avi purified from the seam cells. In each of the DLG-1 and CDC-42 purifications, the highest number of identified *C. elegans* peptides matched the bait protein.

For DLG-1, we again identified the two known interacting proteins AJM-1 (with 1 peptide) and ATAD-3 (with 3 peptides), demonstrating that these protein partners can also be identified from a single tissue (Table 2). Analysis of CDC-42 purified specifically from the seam cells led to the identification of RHI-1, a Rho GDI that had previously been described to interact with CDC-42 in *C. elegans*.[45] RHI-1 was found in 2 out of 3 samples with an average peptide count of 7, and was not identified in any of the control samples (Table 3). Taken together, these data demonstrate that our approach based on the *in vivo* biotinylation of proteins of interest is able to purify proteins from specific tissues, and is able to co-purify known protein interaction partners. Currently, little is known about the function of LET-413 and LGL-1. Hence, detailed analysis of the composition of all five polarity protein complexes is likely to reveal novel insight in the regulation of apical-basal polarity and tissue-specific differences in polarity complexes.

Table 2. Mass spectrometry results of intestine-specific purification of DLG-1.

Protein	# PSMs
DLG-1::GFP::Avi	278
ATAD-3	2
AJM-1	1

Table 3. Mass spectrometry results of seam cell-specific purification of CDC-42 and control construct in triplo.

Protein	# PSMs control 1	# PSMs control 2	# PSMs control 3	# PSMs sample 1	# PSMs sample 2	# PSMs sample 3	In x/x controls	In x/x samples
Avi::GFP::CDC-42	-	1	1	34	20	8	2/3	3/3
RHI-1	-	-	-	9	4	-	0/3	2/3

DISCUSSION

Here, we developed an *in vivo* biotinylation-based AP/MS approach to purify protein complexes from specific tissues in *C. elegans*. Biotinylation only occurs in cells where both an Avi-tagged bait protein and the bacterial biotin ligase BirA are present, and biotinylation in a single tissue is accomplished by expressing BirA from a tissue-specific promoter. Experiments where Avi-tagged protein and BirA are expressed in two adjacent tissues confirmed that biotinylation and purification are indeed highly tissue-specific. The tag we developed includes the Avi sequence as well as GFP, to visualize the subcellular localization of tagged proteins. *In vivo* imaging of 5 different polarity regulators tagged with GFP::Avi demonstrated that the proteins were localized to their appropriate cortical domains. Complementation assays performed with two of the transgenes we constructed showed that the GFP::Avi tag does not interfere with protein function. Finally, analysis of DLG-1 purified from the intestine and CDC-42 purified from the seam cells identified several known interacting proteins, confirming that our approach is indeed able to identify *bona fide* protein interactions from specific tissues.

Our approach should be widely applicable for the identification of protein complexes for many proteins, provided that the bait protein is soluble after lysis and can be tagged at either the N-terminus or C-terminus without interfering with its function. In addition, our approach can be used to identify tissue-specific protein modifications of the bait protein, such as phosphorylations, and to reveal differences in protein complex composition over time by using synchronized cultures. Two other methods using BirA-mediated *in vivo* biotinylation have been described in *C. elegans*. One study used *in vivo* biotinylation of the Histone H3.3 protein for the purification of chromatin and epigenetic profiling.[46] Another study used biotinylation of the nuclear pore complex component NPP-9 to purify nuclei from specific tissues.[47] Thus, *in vivo* biotinylation is proving to be a versatile method for tissue-specific biochemical approaches in *C. elegans*.

We are currently analyzing all of the remaining samples by mass spectrometry. Although we have clearly demonstrated the ability to identify relevant interacting proteins from specific tissues, an actual tissue-specific protein interaction remains to be identified. A comparison between the interaction partners of DLG-1 in epithelia and neuronal cells is the most likely to yield tissue-specific differences in complex composition. In addition to its role in epithelial polarity, the mammalian DLG-1 homolog PSD95 functions both in epithelia and in the postsynaptic densities of neurons, where it interacts with a number of neuron-

specific proteins including several synaptic adhesion molecules, neurotransmitters and ion channels.[48,49]

From the limited number of samples we have analyzed thus far it is already clear that a large number of contaminant proteins will be frequently identified, including histone and ribosome subunits. The inclusion of controls expressing only the GFP::Avi tag and performing three replicate experiments will be crucial to differentiate between true interactors and contaminants. For CDC-42, using the stringent requirement of considering only proteins identified in at least two bait samples but not in any of the control samples, the known interactor RHI-1 was the highest scoring protein identified next to the bait itself. This indicates that our experimental design will have sufficient differentiating power to identify relevant interactions.

In the approach we present here, we added the GFP::Avi tag to a fosmid carrying the gene of interest. For most genes, a fosmid is available that likely encompasses all of the regulatory sequences, including large regions upstream of the predicted translational start codon and downstream of the stop codon. Thus, this approach is generally applicable and should yield accurate expression patterns for most genes. Recently however, we and others developed CRISPR/Cas9-based genome engineering for *C. elegans*, which for the first time enables the endogenous tagging of any *C. elegans* gene of interest.[50] An attractive future improvement therefore is to incorporate the GFP::Avi tag at the endogenous locus. This has a number of advantages. First, all regulatory sequences will be present, even if they extend beyond the normal size limit of fosmids. Second, expression levels should mimic endogenous levels, as no extra copies of the gene of interest will be present. Third, insertions generated through homologous recombineering are compatible with germline and early embryonic expression.[50] Fourth, if inactivation of the gene of interest normally causes a visible phenotype, functionality of the tagged protein is automatically tested since no untagged protein will be produced. Finally, the time needed to construct a complete strain should be significantly reduced, as a number of cloning steps are eliminated and there is no need for crosses with a mutant background.

The expression patterns of the GFP::Avi tagged proteins we observed confirmed known localizations, but also yielded a number of novel observations. First, we observed expression of DLG-1::GFP::Avi in the ventral nerve cord, which is consistent with the neuronal function of the mammalian homolog PSD95, but had not yet been observed in *C. elegans*. Second, we observed no expression of PAR-3::GFP::Avi in the intestine, which is surprising given the broad role of PAR-3 in epithelial tissues. One explanation is that PAR-3 is indeed absent or only present at low levels in the intestine during the larval stages. Studies of the expression pattern of PAR-3 in the intestine during embryonic development showed a peak of PAR-3 protein levels during polarization, indicating that the requirement of PAR-3 may be reduced when cells are fully polarized.[29] This would fit with our observed lack of expression in the intestine during larval development. Alternatively, PAR-3::GFP::Avi may not be expressed in the intestine for technical reasons. For example, the regulatory sequences of PAR-3 may extend beyond the region included in the fosmid we used. Alternatively, the

par-3 splice variants we tagged may not be expressed in the intestine. A large and increasing number of splice variants are predicted for *par-3* (14 in Wormbase version WS241), and these include at least two alternative start and two alternative stop codons. The two alternative start sites have already been documented to affect the expression pattern of *par-3*, and it is conceivable that different 3'-ends can also affect *par-3* expression.[51] Finally, we noticed that PAR-3::GFP::Avi localized apically in the seam cells. In *Drosophila* and mammalian epithelial cells, PAR-3 localizes sub-apically, distinct from the apical localization pattern of PAR-6 and aPKC.[19,21,40] Similarly, in the *C. elegans* embryonic intestine, PAR-3 colocalizes with apical junction markers.[29] In the seam cells however, we observed localization of PAR-3 at the entire apical domain. It will be important to exclude the possibility that the integrated array expresses PAR-3::GFP::Avi at higher than endogenous levels, which could overwhelm the normal mechanisms that relocate PAR-3 to the sub-apical region. In future experiments we will therefore examine the localization of PAR-3 using an endogenous fusion generated using CRISPR/Cas9.

ACKNOWLEDGEMENTS

This work is supported by Innovational Research Incentives Scheme Vidi grant 864.09.008, financed by the Netherlands Organization for Scientific Research (NWO/ALW). Part of this research was performed within the framework of the PRIME-XS project, grant number 262067, funded by the European Union 7th Framework Program and the Netherlands Organization for Scientific Research (NWO) supported large scale proteomics facility Proteins@Work (project 184.032.201). Some strains were provided by the CGC, which is funded by NIH Office of Research Infrastructure Programs (P40 OD010440).

MATERIAL AND METHODS

Cloning of GFP::Avi tags

Standard molecular cloning procedures were used to generate the N- and C-terminal GFP::Avi tag constructs pMB71 and pMB72, shown in Figure 1. TEV and Avi sequences were obtained as synthetic DNA constructs from GenScript (http://www.genscript.com). Codon-optimized GFP sequences and GalK sequences used were from vector pBALU1.[37] The vector backbone used was pUC19. Full sequences of the final constructs are available upon request.

Recombineering

The recombineering approach we used has been described previously.[37] Briefly, the fosmid to be engineered is transformed into *E. coli* strain SW105, a *galK* defective strain carrying a heat shock inducible λ Red recombinase, and an arabinose inducible Flp recombinase. Bacteria carrying the fosmid are transformed with a PCR product consisting of the tag to be introduced, which also carries the wild-type *galK* sequence, flanked on both sides by 50 nt identical to the insertion site. Expression of λ Red induces homologous recombination between the fosmid and the PCR product, and successful recombinants are selected on media with galactose as the only carbon source. Finally, the *galK* sequence, which is flanked by FRT sites, is eliminated from the fosmid by expression of the Flp recombinase.

We used the following fosmids and primers to generate the PCR products used as templates for homologous recombination: for tagging *dlg-1* we used the C-terminal tag, fosmid WRM067dB05

and primers rec_dlg-1_C-term_F (actccatcatcagccgtgaatcgcagacgccaatttgggtgccacgtcatggaggaggatctggaggaggaggatctggaggagga) and rec_dlg-1_C-term_R (acatatttcttgaagaaacgattatttgtctaaaaaatatccaatttcatctattcatgccattcaatcttctgagcttcg), for *let-413* the C-terminal tag, fosmid WRM0640dF02 and primers rec_let-413_C-term_F (ggtccccatcgccagtttcgagaacatctgtgagtaggccatgtgagtatggaggaggatctggaggaggaggatctggaggagga) and rec_let-413_C-term_R (gaatgtcaaaaaaaaaacgtctaatgtctagttttcagccaaaatcggcctcattcatgccattcaatcttctgagcttcg), for *lgl-1* the C-terminal tag, fosmid WRM065bB11 and primers rec_lgl-1_C-term_F (gaagtacggtgaattt gaacttcgcggttggagcagtacgcacaagtcaggaggaggatctggaggaggaggatctggaggagga) and rec_lgl-1_C-term_R (aaaattaatatatatcaacaggaaaacgatttttaaaaaaaatgcatctattcatgccattcaatcttctgagcttcg), for *par-3* the C-terminal tag, fosmid WRM064bG02 and primers rec_par-3_C-term_F (gccaataccgtcgcaga gatcagggaccgcctcatcgttttccccagtacggaggaggatctggaggaggaggatctggaggagga) and rec_par-3_C-term_R (gattccgtatttttcgcggctgcgtaatataactttgagaaaaaactgacctattcatgccattcaatcttctgagcttcg), for *cdc-42* the N-terminal tag, fosmid WRM0612bG08 and primers rec_cdc-42_N-term_F (ctataaagacgtaattttaatacttttattcatttttttttttcaggcgaaaaaaatgggacttaatgatattttcgaagctcag) and rec_cdc-42_N-term_R (gttttaccgacagctccatctccaacgacgacgcacttgatcgtctgcatacctcctcctccagatcctcctcct). To check whether the recombineering procedure was successful we performed a PCR reaction on the constructs with primers rec_dlg-1_C-term_check_F (aagctcaagcgcagtattcc) and rec_dlg-1_C-term_check_R (tttcttgaattgagaacttggaaa) for *dlg-1*, primers rec_let-413_C-term_check_F (cgattggtattccgattggt) and rec_let-413_C-term_check_R (gccgaacagtaacggagatt) for *let-413*, primers rec_lgl-1_C-term_check_F (gggagttatgtacaggcatctagta) and rec_lgl-1_C-term_check_R (taagccagccgctagcac) for *lgl-1*, primers rec_par-3_C-term_check_F (tatgccgcgaaggagaagta) and rec_par-3_C-term_check_R (ttcgctcagcggaattatc) for *par-3*, and primers rec_cdc-42_N-term_check_F (tcgtttattaaggcgtttaccg) and rec_cdc-42_N-term_check_R (cgatcgtaatcttcctgtcc) for *cdc-42*.

Cloning the BirA constructs

Standard molecular cloning techniques were used to generate the BirA expression constructs pMB73 (*Pelt-2::BirA*), pMB74 (*Pwrt-2::BirA*), pMB75 (*Prps-27::BirA*), and pBT331 (*Prgef-1::BirA*). pMB73-75 were generated in the pBluescript SK- backbone, and pBT331 was generated in the pPD95 backbone. The *C. elegans* codon optimized BirA sequence was obtained as a synthetic DNA construct from GenScript. The 3'-UTRs and promoters used were PCR amplified from *C. elegans* genomic DNA. pMB73-75 contain 108 bp of the *let-858* 3'-UTR, and pBT331 contains 865 bp of the *unc-32* 3'-UTR. Promoter lengths used were 3639 bp for *Pelt-2*, 1640 bp for *Pwrt-2*, 837 bp for *Prps-27*, and 3469 bp for *Prgef-1*. Full construct sequences are available upon request.

Cloning the control constructs

Standard molecular cloning techniques were used to generate expression constructs pMB76 (*Pelt-2::GFP::Avi*) and pMB77 (*Pwrt-2::GFP::Avi*). Both vectors contain *C. elegans* codon-optimized GFP cloned from vector pBALU1 after removal of *GalK* sequences by expression of Flp recombinase, 2xTEV::Avi sequences obtained as synthetic DNA from GenScript, and 108 bp of the *let-858* 3'-UTR.[37] The 3'-UTR and the promoters used were PCR amplified from *C. elegans* genomic DNA. Promoter lengths used were 3654 bp for *Pelt-2* and 1640 bp for *Pwrt-2*. The backbone used was pBluescript SK-. Full construct sequences are available upon request.

***C. elegans* strains and culture conditions**

Unless otherwise indicated, strains were maintained at 15°C as previously described.[52] Transgenic strains were generated by injecting constructs into the gonad of young adult animals using

conventional micro-injection procedures. The amounts of each construct injected are indicated for each strain, and were supplemented to a final DNA concentration of 80 ng/µl with *Pst*I digested phage λ DNA. The resulting transgenic strains carrying extrachromosomal arrays were subjected to gamma irradiation to integrate the construct into the *C. elegans* genome. The following strains were generated:

BOX20: *mibIs7[Pwrt-2::BirA 10 ng/µl + Pmyo-2::mCherry 2,5 ng/µl]*II

BOX27: *mibIs14[Pelt-2::BirA 10 ng/µl + Pmyo-2::mCherry 2,5 ng/µl]*I

BOX41: *mibIs23[lgl-1::GFP-2xTEV-Avi 10 ng/µl, Pmyo-3::mCherry 5 ng/µl]*V

BOX43: *mibIs25[Avi-2xTEV-GFP::cdc-42 10 ng/µl, Pmyo-3::mCherry 5 ng/µl]*X

BOX51: *mibIs26[par-3::GFP-2xTEV-Avi 10 ng/µl, Pmyo-3::mCherry 5 ng/µl]*V

BOX55: *mibIs30[let-413::GFP-2xTEV-Avi 10 ng/µl, Pmyo-3::mCherry 5 ng/µl]*X

BOX56: *mibIs31[dlg-1::GFP-2xTEV-Avi 10 ng/µl, Pmyo-3::mCherry 5 ng/µl]*V

BOX58: *mibIs33[Prps-27::BirA 10 ng/µl, Pmyo-2::mCherry 2,5 ng/µl]*I

BOX61: *mibIs36[Pwrt-2::GFP-2xTEV-Avi 10 ng/µl, Prab-3::mCherry 5 ng/µl]*X

BOX62: *mibIs37[Pelt-2::GFP-2xTEV-Avi 10 ng/µl, Prab-3::mCherry 5 ng/µl]*X

BOX65: *mibIs40[Prps-27::GFP-2xTEV-Avi 10 ng/µl, Prab-3::mCherry 5 ng/µl]*III

BOX99: *mibIs14[Pelt-2::BirA 10 ng/µl, Pmyo-2::mCherry 2,5 ng/µl]*I; *mibIs37[Pelt-2::GFP-2xTEV-Avi 10 ng/µl, Prab-3::mCherry 5 ng/µl]*X

BOX100: *mibIs7[Pwrt-2::BirA 10 ng/µl, Pmyo-2::mCherry 2,5 ng/µl]*II; *mibIs37[Pelt-2::GFP-2xTEV-Avi 10 ng/µl, Prab-3::mCherry 5 ng/µl]*X

BOX101: *mibIs14[Pelt-2::BirA 10 ng/µl, Pmyo-2::mCherry 2,5 ng/µl]*I; *mibIs36[Pwrt-2::GFP-2xTEV-Avi 10 ng/µl, Prab-3::mCherry 5 ng/µl]*X

BOX102: *mibIs7[Pwrt-2::BirA 10 ng/µl, Pmyo-2::mCherry 2,5 ng/µl]*II; *mibIs36[Pwrt-2::GFP-2xTEV-Avi 10 ng/µl, Prab-3::mCherry 5 ng/µl]*X

BOX103: *mibIs26[par-3::GFP-2xTEV-Avi 10 ng/µl, Pmyo-3::mCherry 5 ng/µl]*V; *mibIs33[Prps-27::BirA 10 ng/µl, Pmyo-2::mCherry 2,5 ng/µl]*I

BOX104: *mibIs26[par-3::GFP-2xTEV-Avi 10 ng/µl, Pmyo-3::mCherry 5 ng/µl]*V; *mibIs7[Pwrt-2::BirA 10 ng/µl, Pmyo-2::mCherry 2,5 ng/µl]*II

BOX105: *mibIs26[par-3::GFP-2xTEV-Avi 10 ng/µl, Pmyo-3::mCherry 5 ng/µl]*V; *mibIs14[Pelt-2::BirA 10 ng/µl, Pmyo-2::mCherry 2,5 ng/µl]*I

BOX106: *mibIs31[dlg-1::GFP-2xTEV-Avi 10 ng/µl, Pmyo-3::mCherry 5 ng/µl]*V; *mibIs33[Prps-27::BirA 10 ng/µl, Pmyo-2::mCherry 2,5 ng/µl]*I; *dlg-1(ok318)*X

BOX107: *mibIs31[dlg-1::GFP-2xTEV-Avi 10 ng/µl, Pmyo-3::mCherry 5 ng/µl]*V; *mibIs7[Pwrt-2::BirA 10 ng/µl, Pmyo-2::mCherry 2,5 ng/µl]*II; *dlg-1(ok318)*X

BOX108: *mibIs31[dlg-1::GFP-2xTEV-Avi 10 ng/µl, Pmyo-3::mCherry 5 ng/µl]*V; *mibIs14[Pelt-2::BirA 10 ng/µl, Pmyo-2::mCherry 2,5 ng/µl]*I; *dlg-1(ok318)*X

BOX110: *mibIs30[let-413::GFP-2xTEV-Avi 10 ng/µl, Pmyo-3::mCherry 5 ng/µl]*X; *mibIs33[Prps-27::BirA 10 ng/µl, Pmyo-2::mCherry 2,5 ng/µl]*I; *dlg-1(ok318)*X

BOX111: *mibIs30[let-413::GFP-2xTEV-Avi 10 ng/µl, Pmyo-3::mCherry 5 ng/µl]*X; *mibIs7[Pwrt-2::BirA 10 ng/µl, Pmyo-2::mCherry 2,5 ng/µl]*II; *dlg-1(ok318)*X

BOX112: *mibIs30[let-413::GFP-2xTEV-Avi 10 ng/µl, Pmyo-3::mCherry 5 ng/µl]*X; *mibIs14[Pelt-2::BirA 10 ng/µl, Pmyo-2::mCherry 2,5 ng/µl]*I; *dlg-1(ok318)*X

BOX113: *mibIs23[lgl-1::GFP-2xTEV-Avi 10 ng/µl, Pmyo-3::mCherry 5 ng/µl]*V; *mibIs33[Prps-27::BirA 10 ng/µl, Pmyo-2::mCherry 2,5 ng/µl]*I; *lgl-1(tm2616)*X

BOX114: *mibIs23[lgl-1::GFP-2xTEV-Avi 10 ng/µl, Pmyo-3::mCherry 5 ng/µl]*V; *mibIs7[Pwrt-2::BirA 10 ng/µl, Pmyo-2::mCherry 2,5 ng/µl]*II; *lgl-1(tm2616)*X

BOX115: *mibIs23[lgl-1::GFP-2xTEV-Avi 10 ng/µl, Pmyo-3::mCherry 5 ng/µl]*V; *mibIs14[Pelt-2::BirA 10 ng/µl, Pmyo-2::mCherry 2,5 ng/µl]*I; *lgl-1(tm2616)*X

BOX116: *mibIs40[Prps-27::GFP-2xTEV-Avi 10 ng/µl, Prab-3::mCherry 5 ng/µl]*III; *mibIs33[Prps-27::BirA 10 ng/µl, Pmyo-2::mCherry 2,5 ng/µl]*I

BOX117: *mibIs40[Prps-27::GFP-2xTEV-Avi 10 ng/µl, Prab-3::mCherry 5 ng/µl]*III; *mibIs7[Pwrt-2::BirA 10 ng/µl, Pmyo-2::mCherry 2,5 ng/µl]*II

BOX118: *mibIs40[Prps-27::GFP-2xTEV-Avi 10 ng/µl, Prab-3::mCherry 5 ng/µl]*III; *mibIs14[Pelt-2::BirA 10 ng/µl, Pmyo-2::mCherry 2,5 ng/µl]*I

Western blot analysis

Protein samples were separated on 10% acrylamide gels, and subjected to western blotting on polyvinylidene difluoride membrane (Immobilon-P; Millipore). Blots were blocked with 5% skim milk in PBST (7 mM Na_2HPO_4, 3 mM NaH_2PO_4, 140 mM NaCl, 5 mM KCl, 0.05% Tween-20) for 1 hour at room temperature. For protein detection, blots were incubated with rabbit polyclonal anti-GFP (Abcam ab6556, 1:1000) or anti-biotin (Abcam ab1227, 1:1000) in PBST + 5% skim milk for 1 hour at room temperature, washed with PBST three times for 10 minutes at room temperature, incubated with anti-rabbit IgG antibody conjugated to horseradish peroxidase (Jackson Immuno Research 111035003, 1:10,000) for 45 minutes at room temperature, washed with PBST three times for 10 minutes at room temperature, and finally washed once with PBS at room temperature for 10 minutes. Blots were developed using enhanced chemiluminescent Western blotting substrate (BioRad).

***C. elegans* liquid culture**

Liquid cultures were started with semi-synchronized L1 animals obtained by starvation. Depending on the growth rate of the transgenic strain to be cultured, 20-60 9 cm NGM plates with OP50 bacteria were seeded with 15-45 L4 animals per plate. After 6 or 7 days at 20°C no bacteria were left on the plates and the plates were covered with starved L1 animals. All animals were washed off the plates and transferred to a 2L Erlenmeyer flask containing 500 ml S-medium supplemented with Penicillin-Streptomycin (5,000 U/mL, Life Technologies 15070-63) diluted 1:100, and Nystatin Suspension (10,000 U/mL, Sigma N1638) diluted 1:1000.[53] A pellet of OP50 *E. coli* bacteria obtained from a 0.5 L overnight culture in lysogeny broth (LB) was added as food source. Animals were allowed to develop until the L3/L4 stage in an incubator at 20°C shaking at 200 rpm. To harvest the animals, the culture was transferred to 50 ml conical tubes and cooled on ice for 20 minutes. Animals were then pelleted by centrifugation. All centrifugation steps in this protocol were performed at 400 g for 2 minutes at 4°C, pooled in a single 50 ml tube, and washed twice in ice-cold M9 lacking $MgSO_4$.[53] Any remaining contaminants were removed through a sucrose float procedure. After the second wash step, animals were resuspended in 20 ml of ice-cold M9 lacking $MgSO_4$, followed by the addition of 20 ml of ice-cold 60% sucrose in H_2O. After vigorous mixing of the sucrose/worm mixture, 4 ml of ice-cold M9 lacking $MgSO_4$ was gently layered on top, and the worms were centrifuged at 400 g for 2 minutes at 4°C. A layer

of animals should now be visible on top of the sucrose, while contaminants have sedimented at the bottom. The sucrose float steps should be performed as quickly as possible or the layer will fail to form properly. To maximize recovery, 30 ml of supernatant was aspirated from the sucrose float, and distributed into four 50 ml tubes which were subsequently filled-up by addition of room-temperature M9 lacking $MgSO_4$. The warm M9 allows the animals to become mobile again and empty their intestines of OP50 bacteria. The 4 tubes were placed on ice to cool down for 30 minutes, after which the animals were washed twice in lysis buffer (150 mM NaCl, 20 mM Tris pH 7.8, 0.5 mM EDTA). During the first wash the animals were again pooled in 1 tube. After a final wash in lysis buffer supplemented with 1% Triton X-100, as much lysis buffer as possible was removed, and the *C. elegans* pellet was frozen in liquid nitrogen and stored at -80°C.

Lysis

C. elegans pellets were lysed using a French pressure cell. The frozen pellets were ground in liquid nitrogen with a mortar and pestle. Per 1 ml of ground pellet, 5 ml of lysis buffer (150 mM NaCl, 20 mM Tris pH 7.8, 0.5 mM EDTA) supplemented with 1% Triton X-100, 0.5 tablet of protease inhibitor (Roche, 05892791001), and 7 µl of β-mercapto-ethanol was added. The suspension was passed through a French press three times. To remove cellular debris, lysates were distributed to 2 ml Eppendorf tubes, centrifuged at 16,000 g for 15 minutes at 4°C, and collected in a fresh conical bottom 15 ml polypropylene tube. The concentration of protein in the lysates was determined by Bradford assay (Bio-Rad) and the lysates were diluted to 1 mg protein/1 ml lysate.

Affinity purification

Affinity purifications were performed with streptavidin-coated beads (Chromotek, HP57.1). Prior to use, beads were washed twice in lysis buffer (150 mM NaCl, 20 mM Tris pH 7.8, 0.5 mM EDTA) supplemented with 1% Triton X-100 in a 1.5 ml Eppendorf tube, pelleting the beads by centrifugation at 7,500 g for 30 seconds. After the final wash, beads were resuspended in lysis buffer, in the same volume as originally taken to be washed. To 15 ml of lysate (15 mg protein) in a 15 ml conical bottom polypropylene tube, 25 µl of beads were added, after which the tubes were rotated at 4°C for 1.5 hours. Following the incubation, beads were pelleted by centrifugation at 3220 g for 5 minutes at 4°C, resuspended in 0.5 ml TEV buffer (20 mM Tris pH 8.0, 150 mM NaCl, and 0.3% NP40), and transferred to an Eppendorf tube. Beads were then washed three times with 500 µl TEV buffer (20 mM Tris pH 8.0, 150 mM NaCl, and 0.3% NP40), pelleting the beads by centrifugation at 7,500 g for 30 seconds at 4°C. Finally, we resuspended the beads in 15 µl TEV buffer and added 2 µl of TEV protease (Promega, V6101) to the samples. We performed TEV cleavage overnight at 4°C in a shaking block for Eppendorf tubes, shaking at 800 rpm.

Mass spectrometry

After overnight TEV cleavage the samples were centrifuged and the supernatant was subjected to mass spectrometry analysis. To remove any residual detergent from the samples the proteins were digested with trypsin using the previously described FASP protocol.[54] Then, resulting peptides were analyzed by nanoflow LC-MS/MS by coupling an Agilent 1200 HPLC system (Agilent Technologies) to an LTQ-Orbitrap Velos mass spectrometer (Thermo Electron, Bremen, Germany) as described previously.[55] Samples were dried, reconstituted in 10% formic acid and delivered to a trap column (AquaTM C18, 5 µm (Phenomenex, Torrance, CA); 20 mm × 100-µm inner diameter, packed in-house) at 5 µl/min in 100% solvent A (0.1 M acetic acid in water). Next, peptides eluted from the trap column onto an analytical column (ReproSil-Pur C18-AQ, 3 µm (Dr Maisch GmbH, Ammerbuch, Germany); 40 cm × 50-µm inner diameter, packed in-house)

at 100 nl/min in a 90-minute or 3-hour gradient from 0 to 40% solvent B (0.1 M acetic acid in 8:2 (v/v) acetonitrile/water) depending on sample amount and complexity. Eluted peptides were introduced by ESI into the mass spectrometer that was operated in a data-dependent acquisition mode. After the survey scan (30 000 FHMW), the 10 most intense precursor ions were selected for subsequent fragmentation in a data-dependent decision tree as described before using HCD (essentially beam type CID), ETD-IT and ETD-FT activation techniques.[55] In brief, doubly charged peptides were subjected to HCD fragmentation and higher charged peptides were fragmented using ETD. The normalized collision energy for HCD was set to 35%. ETD was enabled with supplemental activation and the reaction time was set to 50 ms for doubly charged precursors.

Raw files were processed with Proteome Discoverer (Beta version 1.3, Thermo). Peptide identification was carried out with Mascot 2.3 (Matrix Science) against a concatenated forward-decoy *C. elegans* database. The following parameters were used: 50 p.p.m. precursor mass tolerance, 0.6 Da fragment ion tolerance for ETD-IT, and 0.02 Da for HCD and ETD-FT modes. Up to two missed cleavages were accepted, carbamidomethylation of cysteines was set up as fixed modification whereas methionine oxidation as variable modification. Mascot results were filtered afterwards with 10 p.p.m. precursor mass tolerance, Mascot Ion Score >20 and a minimum of 7 residues per peptide. Using these criteria, FDRs were calculated to be below 1%.

Microscopy and image processing

Imaging was performed on a spinning disc confocal sytem, consisting of a Nikon Ti-U inverted microscope with a motorized stage and a Piezo Z stage, and a PLAN APO VC 60X oil objective; a Yokogawa CSU-X1 spinning disk unit, equipped with a dual dichroic mirror set for laser wavelengths 488 nm and 561 nm; 488 nm and 561 nm solid state 50 mW lasers, controlled by an Andor revolution 500 series AOTF Laser modulator and combiner; Semrock 525 and 617 nm single band fluorescence emission filters (30 and 73 nm bandwith respectively); Semrock 525 single band fluorescence filter (center wavelength of 525 nm, with a GMBW of 30 nm); and an Andor iXON DU-885 monochrome EMCCD+ camera. All imaging was done using Andor iQ imaging software version 1.1. Maximum projections were generated from a series of slices of a Z-stack with ImageJ and processed with Adobe Photoshop CS6 and Adobe Illustrator CS6.

REFERENCES

1. Altelaar, M. A., Munoz, J. & Heck, A. J. Next-generation proteomics: towards an integrative view of proteome dynamics. Nat. Rev. Genet. 14, 35–48 (2013).
2. Gavin, A.-C. *et al*. Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature 415, 141–7 (2002).
3. Butland, G. *et al*. Interaction network containing conserved and essential protein complexes in *Escherichia coli*. Nature 433, 531–7 (2005).
4. Krogan, N. J. *et al*. Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. Nature 440, 637–43 (2006).
5. Guruharsha, K. G. *et al*. *Drosophila* Protein interaction Map (DPiM). Fly (Austin). 6, 246–53 (2012).
6. Desai, A. *et al*. KNL-1 directs assembly of the microtubule-binding interface of the kinetochore in *C. elegans*. Genes Dev. 17, 2421–35 (2003).
7. Cheeseman, I. M. *et al*. A conserved protein network controls assembly of the outer kinetochore and its ability to sustain tension. Genes Dev. 18, 2255–68 (2004).
8. Gottschalk, A. *et al*. Identification and characterization of novel nicotinic receptor-associated proteins in *Caenorhabditis elegans*. EMBO J. 24, 2566–78 (2005).

9. Ding, L., Spencer, A., Morita, K. & Han, M. The developmental timing regulator AIN-1 interacts with miRISCs and may target the argonaute protein ALG-1 to cytoplasmic P bodies in *C. elegans*. Mol. Cell 19, 437–47 (2005).
10. Duchaine, T. F. *et al*. Functional proteomics reveals the biochemical niche of *C. elegans* DCR-1 in multiple small-RNA-mediated pathways. Cell 124, 343–54 (2006).
11. Zhang, L. *et al*. Systematic identification of *C. elegans* miRISC proteins, miRNAs, and mRNA targets by their interactions with GW182 proteins AIN-1 and AIN-2. Mol. Cell 28, 598–613 (2007).
12. Gassmann, R. *et al*. A new mechanism controlling kinetochore-microtubule interactions revealed by comparison of two dynein-targeting components: SPDL-1 and the Rod/Zwilch/Zw10 complex. Genes Dev. 22, 2385–99 (2008).
13. Gu, W. *et al*. Distinct argonaute-mediated 22G-RNA pathways direct genome surveillance in the *C. elegans* germline. Mol. Cell 36, 231–44 (2009).
14. Moresco, J. J., Carvalho, P. C. & Yates, J. R. Identifying components of protein complexes in *C. elegans* using co-immunoprecipitation and mass spectrometry. J. Proteomics 73, 2198–204 (2010).
15. Galli, M. *et al*. aPKC phosphorylates NuMA-related LIN-5 to position the mitotic spindle during asymmetric division. Nat. Cell Biol. 13, 1132–8 (2011).
16. St Johnston, D. & Ahringer, J. Cell polarity in eggs and epithelia: parallels and diversity. Cell 141, 757–74 (2010).
17. McCaffrey, L. M. & Macara, I. G. Epithelial organization, cell polarity and tumorigenesis. Trends Cell Biol. 21, 727–35 (2011).
18. Kemphues, K. J., Priess, J. R., Morton, D. G. & Cheng, N. S. Identification of genes required for cytoplasmic localization in early *C. elegans* embryos. Cell 52, 311–20 (1988).
19. Harris, T. J. C. & Peifer, M. The positioning and segregation of apical cues during epithelial polarity establishment in *Drosophila*. J. Cell Biol. 170, 813–23 (2005).
20. Totong, R., Achilleos, A. & Nance, J. PAR-6 is required for junction formation but not apicobasal polarization in *C. elegans* embryonic epithelial cells. Development 134, 1259–68 (2007).
21. Morais-de-Sá, E., Mirouse, V. & St Johnston, D. aPKC phosphorylation of Bazooka defines the apical/lateral border in *Drosophila* epithelial cells. Cell 141, 509–23 (2010).
22. Walther, R. F. & Pichaud, F. Crumbs/DaPKC-dependent apical exclusion of Bazooka promotes photoreceptor polarity remodeling. Curr. Biol. 20, 1065–74 (2010).
23. Tepass, U., Theres, C. & Knust, E. crumbs encodes an EGF-like protein expressed on apical membranes of *Drosophila* epithelial cells and required for organization of epithelia. Cell 61, 787–99 (1990).
24. Bilder, D. & Perrimon, N. Localization of apical epithelial determinants by the basolateral PDZ protein Scribble. Nature 403, 676–80 (2000).
25. Bilder, D., Li, M. & Perrimon, N. Cooperative regulation of cell polarity and growth by *Drosophila* tumor suppressors. Science. 289, 113–6 (2000).
26. Goehring, N. W. & Grill, S. W. Cell polarity: mechanochemical patterning. Trends Cell Biol. 23, 72–80 (2013).
27. Rodriguez-Boulan, E. & Macara, I. G. Organization and execution of the epithelial polarity programme. Nat. Rev. Mol. Cell Biol. 15, 225–42 (2014).
28. Tepass, U. The apical polarity protein network in *Drosophila* epithelial cells: regulation of polarity, junctions, morphogenesis, cell growth, and survival. Annu. Rev. Cell Dev. Biol. 28, 655–85 (2012).
29. Achilleos, A., Wehman, A. M. & Nance, J. PAR-3 mediates the initial clustering and apical localization of junction and polarity proteins during *C. elegans* intestinal epithelial cell polarization. Development 137, 1833–42 (2010).
30. Shibata, Y., Fujii, T., Dent, J. A., Fujisawa, H. & Takagi, S. EAT-20, a novel transmembrane protein with EGF motifs, is required for efficient feeding in *Caenorhabditis elegans*. Genetics 154, 635–46 (2000).
31. Bossinger, O., Klebes, A., Segbert, C., Theres, C. & Knust, E. Zonula adherens formation in *Caenorhabditis elegans* requires *dlg-1*, the homologue of the *Drosophila* gene discs large. Dev. Biol. 230, 29–42 (2001).

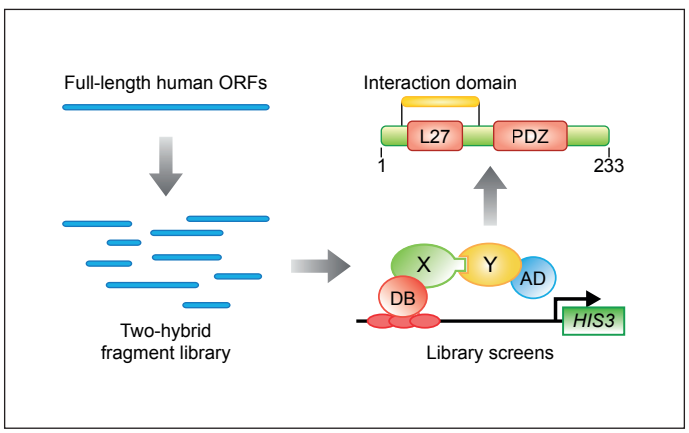
32. Segbert, C., Johnson, K., Theres, C., van Fürden, D. & Bossinger, O. Molecular and functional analysis of apical junction formation in the gut epithelium of *Caenorhabditis elegans*. Dev. Biol. 266, 17–26 (2004).
33. Legouis, R. *et al*. LET-413 is a basolateral protein required for the assembly of adherens junctions in *Caenorhabditis elegans*. Nat. Cell Biol. 2, 415–22 (2000).
34. Hoege, C. *et al*. LGL can partition the cortex of one-cell *Caenorhabditis elegans* embryos into two domains. Curr. Biol. 20, 1296–303 (2010).
35. De Boer, E. *et al*. Efficient biotinylation and single-step purification of tagged transcription factors in mammalian cells and transgenic mice. Proc. Natl. Acad. Sci. U. S. A. 100, 7480–5 (2003).
36. Schatz, P. Use of peptide libraries to map the substrate specificity of a peptide-modifying enzyme: a 13 residue consensus peptide specifies biotinylation in *Escherichia coli*. Biotechnol. (N Y). 10, 1138–43 (1993).
37. Tursun, B., Cochella, L., Carrera, I. & Hobert, O. A toolkit and robust pipeline for the generation of fosmid-based reporter genes in *C. elegans*. PLoS One 4, e4625 (2009).
38. Christner, J. E., Schlesinger, M. J. & Coon, M. J. Enzymatic activation of biotin. J. Biol. Chem. 239, 3997–4005 (1964).
39. Lockwood, C. A., Lynch, A. M. & Hardin, J. Dynamic analysis identifies novel roles for DLG-1 subdomains in AJM-1 recruitment and LET-413-dependent apical focusing. J. Cell Sci. 121, 1477–87 (2008).
40. Izumi, Y. *et al*. An atypical PKC directly associates and colocalizes at the epithelial tight junction with ASIP, a mammalian homologue of *Caenorhabditis elegans* polarity protein PAR-3. J. Cell Biol. 143, 95–106 (1998).
41. Aceto, D., Beers, M. & Kemphues, K. J. Interaction of PAR-6 with CDC-42 is required for maintenance but not establishment of PAR asymmetry in *C. elegans*. Dev. Biol. 299, 386–97 (2006).
42. Palmer, E. & Freeman, T. Investigation into the use of C- and N-terminal GFP fusion proteins for subcellular localization studies using reverse transfection microarrays. Comp. Funct. Genomics 5, 342–53 (2004).
43. Stathakis, D. G., Hoover, K. B., You, Z. & Bryant, P. J. Human postsynaptic density-95 (PSD95): location of the gene (DLG4) and possible function in nonneural as well as in neural tissues. Genomics 44, 71–82 (1997).
44. Gotta, M., Abraham, M. C. & Ahringer, J. CDC-42 controls early cell polarity and spindle orientation in *C. elegans*. Curr. Biol. 11, 482–8 (2001).
45. Yap, S. F., Chen, W. & Lim, L. Molecular characterization of the *Caenorhabditis elegans* Rho. Eur. J. Biochem. 1100, 1090–100 (1999).
46. Ooi, S. L., Henikoff, J. G. & Henikoff, S. A native chromatin purification system for epigenomic profiling in *Caenorhabditis elegans*. Nucleic Acids Res. 38, e26 (2010).
47. Steiner, F. A., Talbert, P. B., Kasinathan, S., Deal, R. B. & Henikoff, S. Cell-type-specific nuclei purification from whole animals for genome-wide expression and chromatin profiling. Genome Res. 22, 766–77 (2012).
48. Han, K. & Kim, E. Synaptic adhesion molecules and PSD-95. Prog. Neurobiol. 84, 263–83 (2008).
49. Kim, E. & Sheng, M. PDZ domain proteins of synapses. Nat. Rev. Neurosci. 5, 771–81 (2004).
50. Dickinson, D. J., Ward, J. D., Reiner, D. J. & Goldstein, B. Engineering the *Caenorhabditis elegans* genome using Cas9-triggered homologous recombination. Nat. Methods 10, 1028–34 (2013).
51. Li, B., Kim, H., Beers, M. & Kemphues, K. Different domains of *C. elegans* PAR-3 are required at different times in development. Dev. Biol. 344, 745–57 (2010).
52. Brenner, S. The genetics of *Caenorhabditis elegans*. Genetics 77, 71–94 (1974).
53. Wood, W. B. The nematode *Caenorhabditis elegans*.
54. Wisniewski, J. R., Zougman, A., Nagaraj, N. & Mann, M. Universal sample preparation method for proteome analysis. Nat. Methods 6, 3–7 (2009).
55. Frese, C. K. *et al*. Improved peptide identification by targeted fragmentation using CID, HCD and ETD on an LTQ-Orbitrap Velos. J. Proteome Res. 10, 2377–88 (2011).

Chapter 2

Identification of human protein interaction domains using an ORFeome-based yeast two-hybrid fragment library

Selma Waaijers[1,2], Thijs Koorman[1,2], Jana Kerver[2] & Mike Boxem[2]

Journal of Proteome Research (2013) Jul 5; 12(7): 3181-92.



[1] Co-first authorship.

[2] Utrecht University, Department of Biology, Padualaan 8, 3584 CH Utrecht, The Netherlands.

ABSTRACT

Physical interactions between proteins are essential for biological processes. Hence, there have been major efforts to elucidate the complete networks of protein-protein interactions, or "interactomes", of various organisms. Detailed descriptions of protein interaction networks should include information on the discrete domains that mediate these interactions, yet most large-scale efforts model interactions between whole proteins only. We previously developed a yeast two-hybrid based strategy to systematically map interaction domains, and generated a domain-based interactome network for 750 proteins involved in *Caenorhabditis elegans* early embryonic development. Here, we expand the concept of Y2H based interaction domain mapping to the genome-wide level. We generated a human fragment library by randomly fragmenting the full-length open reading frames (ORFs) present in the human ORFeome collection. Screens using several proteins required for cell division or polarity establishment as baits demonstrate the ability to accurately identify interaction domains for human proteins using this approach, while the experimental quality of the Y2H data was independently verified in co-affinity purification assays. The library generation strategy can easily be adapted to generate libraries from full-length ORF collections of other organisms.

INTRODUCTION

Eukaryotic proteins are modular in nature: most are predicted to contain multiple domains that can fold independently and retain their specific biological activity when expressed in isolation.[1,2] Domains are thought to represent functional units that facilitate the evolution of proteins with new or modified functions.[3] An important role of protein domains is the mediation of molecular interactions.[4,5] Interaction domains can interact with other interaction domains, with short linear peptide motifs, and with other macromolecules such as DNA or lipids. The specific combination of interaction domains and motifs determines the network of interactions a protein engages in, partly dictating its function.

Large-scale efforts using affinity purification/mass spectrometry or the yeast two-hybrid (Y2H) system have made considerable progress describing the protein interactomes of several model organisms, as well as humans.[6–13] However, large-scale interaction mapping approaches largely disregard the modular nature of proteins, and current domain-domain interaction database entries are either inferred from protein structures in the Protein Data Bank (PDB), or computationally predicted.[14–16] Capturing information on interaction domains is essential for an accurate description of interactome networks, and facilitates generating biological hypotheses from such networks. In addition to generating more accurate descriptions of protein interactions, domain boundaries that are experimentally defined can be a valuable resource for approaches that rely on expressing large amounts of purified protein, such as structure determination by crystallography and high-throughput screens for drug discovery.[5,17,18]

The yeast two-hybrid system is one of the most powerful and widely used methods available to date to identify protein interactions: it is cost-effective, easy to perform, scalable to whole-genome levels, and can be applied to map interactions of any species, including cross-species interactions such as virus-host interactions.[19,20] Since its inception, major improvements have been made to the Y2H system to increase the data quality and effectiveness of the approach.[21–26] We previously demonstrated that the Y2H system can be used to systematically identify protein interaction domains for a set of 750 proteins involved in *C. elegans* early embryogenesis.[27] Here, we expand our approach and demonstrate the use of a genome-wide fragment library to accurately identify interaction domains of human proteins. We generated a human Y2H fragment library by mechanically fragmenting open reading frame (ORF) clones from the human ORFeome 5.1, a collection of 15,483 human ORF clones, and screened the library with a series of bait proteins with well described interactions.[28] For each of 7 interactions where an interaction domain had previously been defined, the minimal binding region identified by our approach matched the published interaction domain. Moreover, independent co-affinity purification assays validated 55% of the interaction domains we identified.

Our experiments demonstrate that interacting regions of proteins can be systematically identified by Y2H from genome-wide ORFeome-derived random fragment libraries. The identification of interaction domains adds additional detail to interactome networks, while the experimentally defined domain boundaries that result in functional proteins in yeast are

a valuable starting point for structure-function analyses, and other approaches requiring expression of protein domains. The pipeline we used to generate the library can be used to generate comparable random fragment libraries for other systems for which ORF collections are available.

RESULTS

ORFeome clones are an ideal resource to generate genome-wide random fragment libraries

To enable the systematic mapping of interaction domains on a genome-wide scale, we set out to generate a genome-wide human AD-fragment library. We had previously used a PCR-based approach to generate a fragment library for 750 *C. elegans* proteins.[27] The attractive feature of this method is that it yields complete control over the contents of the library: each protein can be systematically covered by a series of fragments of the different sizes, each clone will be in frame, and the library will be normalized. However, using PCR to generate a genome wide library is both costly and time consuming, since each fragment has to be amplified individually. As alternative strategies, we considered the use of random-primed cDNA libraries and the random fragmentation of protein coding sequences, both of which have been used to generate Y2H libraries and should be able to identify interaction domains.[29,30] One of our objectives was to generate a library in which all genes are equally represented. Since random-primed cDNA libraries will contain widely varying levels of clones for different genes, dependent on their expression levels, we decided to generate a library by random fragmentation of ORF clones. This approach is made possible by the development of ORFeome projects, which aim to provide complete sets of protein-encoding ORFs for organisms including viruses, bacteria, nematodes and humans.[28,31–36] Using an ORFeome library as the starting material offers several advantages. First, library generation starts with an equal amount of each ORF construct, resulting in a normalized representation of all genes in the library. Second, the library can contain both fragmented and full-length constructs, which increases the chances of identifying a given interaction. Third, library construction can be efficiently accomplished by working with large pools of PCR products generated with primers annealing to vector sequences flanking the ORF. The main drawback of an ORFeome based library is that genes not present in the ORFeome collection will also be absent from the library. The library constructed here is based on version 5.1 of the human ORFeome, which contains 15,483 different ORFs corresponding to 12,794 distinct genes.[28]

Construction of the human AD-fragment library

The strategy used to generate the human AD-fragment library is shown in Figure 1. We PCR amplified each of the 15,483 full-length ORF constructs using a set of universal primers annealing to the vector sequences flanking the Gateway *attL* sites. PCR amplifications were performed in 384 well plates, and 12 samples from each plate were analyzed on a gel to monitor the PCR success rate, which was >95%. The PCR products were pooled into 5 pools based on size (<500 bp, 500 – 1000 bp, 1000 – 1500 bp, 1500 – 2000 bp, >2000 bp) (Fig. 2A). After a short Exonuclease III treatment to remove the Gateway attL tails flanking the

ORFs, the PCR pools were mechanically fragmented using a Covaris S2 ultrasonicator (Fig. 2B). We chose mechanical fragmentation over enzymatic approaches such as frequently cutting restriction enzymes, as this approach fragments DNA without any specific sequence preference. Taking into account that most self-folding protein domains are estimated to be between 100 and 200 amino acid residues long, we aimed for a median fragment size of 700 bp for ORFs > 1000 bp in length, a median fragment size of 500 bp for ORFs 500 – 1000 bp, and a median fragment size of 250 bp for ORFs < 500 bp.[2] Settings for the ultrasonicator

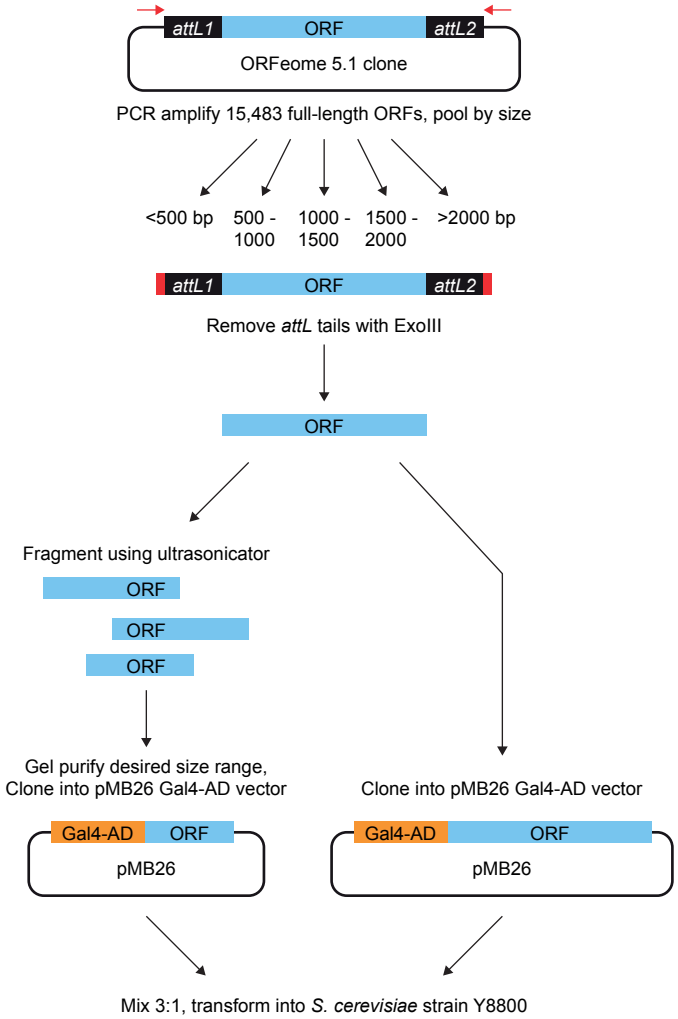


Figure 1. Schematic representation of the pipeline used to generate the human fragment library. Full-length ORFs were PCR amplified from the ORFeome 5.1 resource using primers annealing just outside the Gateway *attL* sequences. After removal of vector sequences by Exonuclease and mechanical fragmentation by ultrasonication, fragments were cloned into a Y2H Gal4-AD vector. A subset of PCR products was cloned without fragmentation to add a full-length component to the library.

were chosen to maximize the yield of fragments in the desired size ranges (Fig. 2C). To further ensure cloning of fragments of the correct size, we gel-purified the desired range from the ulrasonicated PCR pools (Fig. 2D). The resulting pools of fragments were ligated into the pMB26 Gal4-AD Y2H vector and transformed into *Escherichia coli*. From each pool we generated sufficient transformants to have a >99% chance that each ORF is fully represented in the library (see methods). Finally, the 5 pools of cloned fragments were mixed together such that all ORFs are represented equally. The final library contains a total of 1.6 x 10^6 clones. Assuming an 85% cloning success rate (see below) and taking into account that 1/6 of the clones are in frame with the Gal4-AD sequence, this corresponds to 2.3 x 10^5 in frame fragment clones, or an average of 14 clones/ORF.

In addition to the generation of fragments, we cloned each PCR pool without fragmentation, to add a set of full-length clones to the library. We generated a total of 69,000 in frame full length clones, again taking into account the average cloning success rate and 1/6 chance of

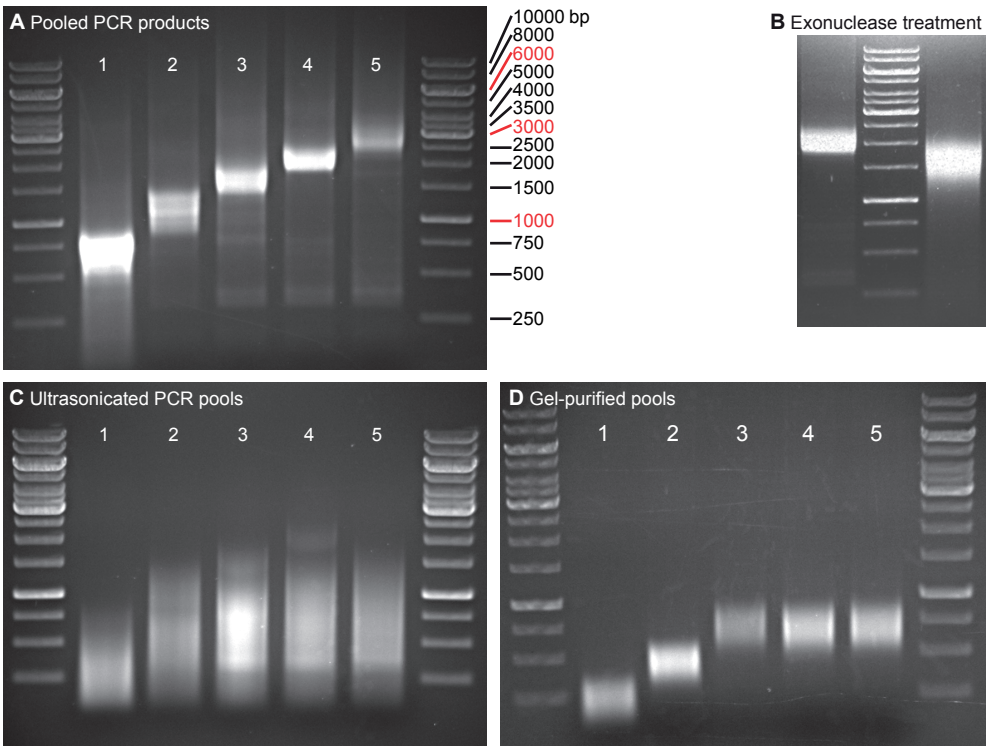


Figure 2. Agarose gel analysis of DNA samples at several steps in the pipeline shown in Figure 1. (A) PCR products collected in 5 pools based on size. (B) Example of the elimination of *attL* tails in a brief exonuclease III treatment. Left: pool 4 prior to treatment, right: pool 4 after exonuclease treatment. (C) DNA pools after ultrasonication on a Covaris S2. (D) DNA pools after gel-purification of desired size range. This represents the final step before ligation into vector pMB26. In all gels, lanes are as follows: 1: pool 1 (ORFs < 500 bp), 2: pool 2 (ORFs 500 – 1000 bp), 3: pool 3 (ORFs 1000 – 1500 bp), 4: pool 4 (ORFs 1500 – 2000 bp), 5: pool 5 (ORFs >2000 bp).

an in frame clone due to the exonuclease treatment. The final human fragment library was made by mixing the fragment and full-length clones in a 3:1 ratio. The library was transformed into *Saccharomyces cerevisiae* strain Y8800 to generate a yeast mating library, with which all screens were performed. The final yeast mating library contains >3 x 10^6 clones.

To verify the cloning success rate and examine the distribution of fragments relative to the full-length ORFs, we PCR amplified and sequenced the insert from 96 randomly selected yeast clones obtained by plating a 1:100.000 dilution of the final mating library on selective plates (SC –Trp). Of these 96 colonies, 15 could not be analyzed due to poor sequence quality. Of the remaining 81 clones, 67 (85%) contained an insert matching a human ORF, of which 14 (21%) were in frame. The average insert length was 400 bp, which is less than the expected median length of 550 bp for the entire library. This indicates either a bias in our random picking of 96 clones, or preferential cloning of a subset of smaller fragments, despite size selection by agarose gel. We next examined the distribution of each fragment

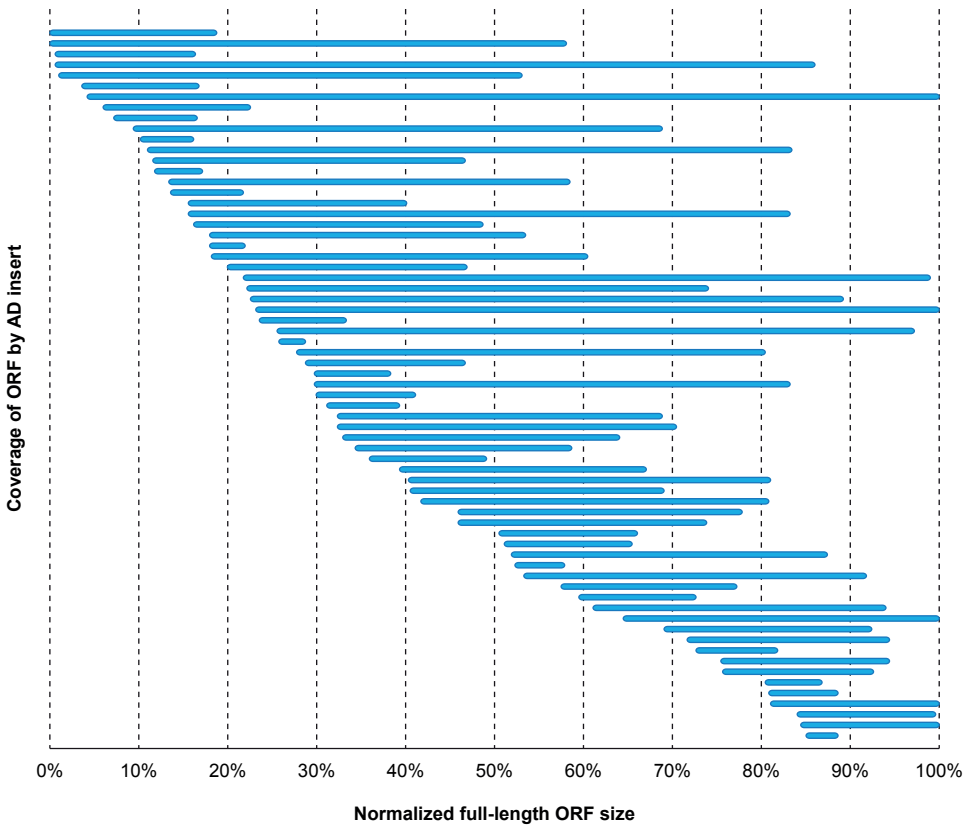


Figure 3. Distribution of a random selection of fragments from the library relative to the full-length ORF size. AD vector inserts were PCR amplified from randomly selected yeast colonies obtained by plating the final Y2H mating library, and analyzed by sequencing. Each blue line represents the area of the corresponding full-length ORF that is covered by the fragment identified from the library.

relative to the full-length ORF, by plotting the region of the full-length ORF covered by the cloned fragment (Fig. 3). The fragments covered from 3% up to 100% of the full-length ORF, with a median of 31%, and the start and end points of the fragments were distributed along the entire length of the ORFs. Together, these results indicate no particular bias in fragmenting of full-length ORFs and subsequent cloning, with the exception of a potential preference for smaller fragments.

Identification of known and novel interactions

To test the human fragment library, we screened the library with 44 bait proteins involved in cell-cycle regulation and cell polarity control (Table 1). For each protein, the corresponding full-length ORF was transferred from the ORFeome entry clone into the pDEST-Gal4-DB vector by Gateway recombinational cloning, and the identity was confirmed by sequencing. Gal4-DB bait clones were then transformed into yeast, and assayed for autoactivation of reporter genes in the absence of an interacting AD-ORF clone. Eight bait strains showed autoactivation of the Y2H reporter genes and were not used further. The remaining 36 bait strains were screened against the human fragment library using a mating-based procedure. From yeast colonies growing on selective plates, we first eliminated the most common source of false-positives: *de novo* autoactivators that arise in the screening process. These are yeast cells that are able to activate reporter genes irrespective of the presence of an interaction. These can efficiently be eliminated by removing all colonies that still activate reporter genes after the AD-ORF plasmid is lost through a counter-selection step based on the sensitivity to cycloheximide conferred by the *CYH2* gene present on the AD plasmid.[21] Next, the AD plasmid inserts were PCR amplified and sequenced to determine the identity and end

Table 1. Bait proteins screened.

Protein	Interactions	Protein	Interactions	Protein	Interactions
ASPM	-	GPSM2	SA	PRKCD	-
CCND3	1	LLGL1	-	PRKCI	2
CDK4	3	LNX1	1	PRKCZ	-
CDKN2B	2	MARK2	-	PROX1	1
CTNNA2	-	MPP4	-	PTEN	-
CTNNA3	4	MPP5	2	RHOA	-
CTNNB1	SA	NUMA1	1	RIC8A	-
DLG2	SA	NUMB	-	RIC8B	SA
DLG5	-	NUMBL	-	STAU1	-
DLGAP5	-	PARD3	-	STAU2	-
DVL2	-	PARD6B	-	STK11	-
DVL3	1	PLK1	SA	VHL	-
GNAI1	-	PLK2	-	YWHAH	SA
GNAO1	-	PLK4	SA	YWHAZ	SA
GPSM1	4	PRKAA2	-		

SA: Self-activating bait

points of the interacting AD-fragments. Sequence traces that were in the wrong orientation or out of frame with the AD coding segment were eliminated. Because it is unlikely that a prey protein is identified twice by chance, only interactions identified in two or more independent yeast colonies were considered valid Y2H interactions. Finally, to ensure that each interaction is reproducible, we retested all interactions in yeast. For every interacting protein pair we isolated a representative Gal4-AD prey plasmid from yeast, and verified the identity by sequencing. Next, the isolated Gal4-AD prey plasmid was transformed into fresh yeast, and mated with the corresponding Gal4-DB bait yeast strain to confirm the interaction.

Together, 11 bait proteins (31%) identified 22 interactions, corresponding to 2 interactors per bait (Table 2). These numbers are comparable to our previous results screening a *C. elegans* fragment library, where 37% of full-length bait proteins identified an average of 2.2 interacting proteins each.[27] To assess the relevance of the interactions we identified, we compared our results with the human interactome HI-2012 (http://interactome.dfci.harvard.edu/H_sapiens/), with interactions in IntAct, and with the literature (Table 2).[37]

Table 2. Protein interactions identified.

			Additional evidence		
DB-X	**AD-Y**	**Hits**	**HI-2012**	**IntAct**	**Literature**
CTNNA3	CTNNB1	77	no	no	yes
CDK4	CDKN2D	34	yes	yes	yes
CDKN2B	PYCRL	30	no	no	no
PROX1	PROX1	26	no	no	no
DVL3	HOMEZ	17	no	no	no
CTNNA3	EHMT2	12	no	no	no
CCND3	CDKN1B	11	no	yes	yes
CDK4	CCND1	8	yes	yes	yes
NUMA1	CCDC57	6	yes	no	no
CTNNA3	JUP	5	no	no	yes
GPSM1	PPP6R3	5	no	no	no
PRKCI	PARD6B	5	yes	yes	yes
CDK4	KLHL32	3	no	no	no
GPSM1	TRIM23	3	no	no	no
LNX1	PRPH	3	no	no	no
MPP5	LIN7A	3	yes	no	yes
CDKN2B	RNF20	2	no	no	no
CTNNA3	SPRY2	2	no	no	no
GPSM1	CBS	2	no	no	no
GPSM1	RALBP1	2	no	no	no
MPP5	ZNF451	2	no	no	no
PRKCI	CRX	2	no	no	no

Of the 22 interactions, 5 were present in HI-2012, 4 were present in IntAct, and 7 were present in the literature. Overall, 8 interactions were supported by these additional sources of evidence, demonstrating that our approach identifies valid interactions.

Identification of minimal interacting regions

For each interaction, the minimal interacting region for an interacting prey protein is defined as the smallest region present in all AD-fragments interacting with a particular bait (Fig. 4 and Fig. S1). Of the 22 interactions we identified, 7 have already been described

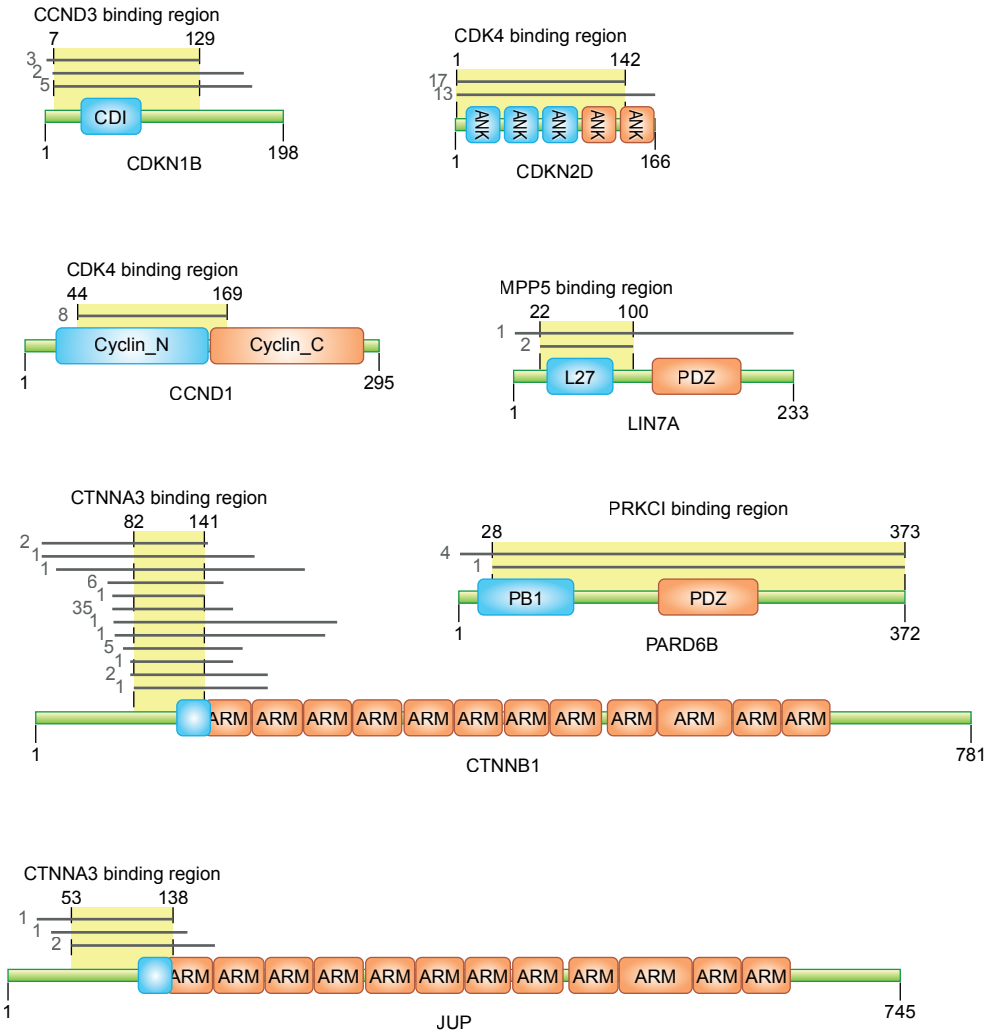


Figure 4. Comparison of minimal interacting regions identified by our Y2H screens (yellow area) with interaction site information present in the literature (blue boxes). Grey lines above the prey protein cartoon indicate distinct prey fragments that we identified, and the numbers before the lines show how often a particular fragment was identified independently in different yeast colonies. Labeled red and blue boxes are SMART domain predictions.

in small-scale literature experiments (Table S1). For each of these 7 interactions, detailed information on the sites that mediate the interaction is known, allowing us to assess the accuracy of the interaction domains we identify. Except for Par6, for which we only found nearly full-length fragments in our screens, we were able to delineate a minimal interacting region for each of the 7 prey proteins involved in the literature-supported interactions (Fig. 4 and Table S2). Importantly, the minimal interacting regions we identified closely match the published interaction sites (compare blue literature interaction sites in Figure 4 with yellow interacting regions we identified). For example, for the homologous proteins β-Catenin and Junction Plakoglobin, we found that a small region upstream of a series of Armadillo repeats mediates their interaction with β-Catenin. These regions overlap the stretch of 29 amino acids previously reported to mediate these interactions.[38] Similarly, we found that CDK4 binds to the N-terminal lobe of Cyclin D1 and to the first 4 Ankyrin repeats of p19Ink4d, that Cyclin D3 binds to the CDI domain of p27Kip1, and that the human Pals1 homolog MPP5 binds to the L27 domain of LIN7. In all cases this is in accordance with the published interaction sites.[39–43]

The smallest interacting region, expressed as a fraction of the full-length protein, was the β-Catenin region that binds to β-Catenin (60 amino acids, or 7.7% of full-length β-Catenin), while the largest interacting region we found was the region of p19Ink4d that mediates binding to CDK4 (142 amino acids, or 85% of full-length p19Ink4d). Thus, both small interaction domains and larger interacting stretches can be accurately identified by our approach.

We also examined novel interacting regions for overlap with known protein domains. As we observed in our previous screens with *C. elegans* proteins, some interacting regions overlap well with predicted domains, while others overlap only partially or match a region of the protein for which no domain predictions are known.[27] Five of the interacting regions we identified correlate well with specific protein domain predictions. The interaction of GPSM1 with TRIM23 is mediated by two B-Box zinc-finger domains, which can be involved in protein interactions (for example in the binding of alpha 4 to the B-Box domain protein MID1).[44,45] GPSM1 also interacts with the C-terminal domain of CBS, which is involved in interacting with the known CBS inhibitor LanCL1.[46] The CDK4 binding site in KLHL32 overlaps with a predicted BTB domain, which is a known protein interaction domain that can mediate heterodimeric interactions.[47] MPP5 binds to two predicted zinc-fingers of the multiple zinc-finger protein ZNF451, and zinc-fingers are known to be able to mediate protein interactions.[48] Finally, α-catenin binds to a region of the methyltransferase EHMT2 encompassing a SET domain with flanking pre- and post-SET domains. SET domains catalyze the methylation of substrate lysines, but together with adjacent pre- and post-SET domains also mediate binding to substrates or other protein binding partners.[49] In summary, the domain predictions overlapping our experimentally defined interacting regions are consistent with a role in mediating a protein-protein interaction.

Independent validation of interactions and interaction domains by co-affinity purification

To further demonstrate the validity of the interactions and interaction domains we identified, we tested each interaction in a co-affinity purification assay. For each interaction, we tested the full-length bait ORF against both the full-length prey ORF and the shortest prey fragment we identified in the Y2H screens. Bait and prey ORFs were transfected into HEK293 cells as Avi-tagged mCherry (Avi-mCherry) and EGFP fusions, respectively, together with the bacterial biotin ligase BirA, which recognizes and biotinylates the Avi tag. Binding of the tagged bait and prey proteins was assessed by affinity purification of biotinylated Avi-mCherry-bait protein using streptavidin-coated beads, followed by detection of co-purified EGFP-prey protein on Western blot. As negative controls, we tested for interaction between Avi-mCherry tagged bait protein and the EGFP tag, as well as between EGFP tagged prey protein and the Avi-mCherry tag. An interaction was considered positive only if no interaction was detected in the negative controls. In total, we were able to reproduce 12 of the 22 protein interactions (55%) in this orthologous assay (Fig. 5, Table 3, Fig. S2). Of these, 6 correspond to interactions previously known from the literature, and 6 are novel interactions. A reproducibility rate of 55% (or 40% if considering only the novel interactions) compares favorably with previous results testing interactions obtained using one method in an orthologous assay. For example, in a study testing a panel of known literature-derived interactions in 5 different protein interaction assays, the maximum reproducibility rate was 36%.[22] Similarly, we previously tested a series of interactions identified by Y2H in the mammalian MAPPIT assay, and obtained a maximum reproducibility rate of 40%.[27] Importantly, of the 12 interactions confirmed by co-affinity purification, 11 were also identified using the fragment prey ORF. This result demonstrates that the minimal interacting regions we identify by Y2H represent valid interaction domains.

DISCUSSION

In this work, we have used physical fragmentation of PCR products generated from an ORFeome resource to generate a human random fragment Y2H library. Using a series of bait proteins involved in cell-cycle regulation of cell polarity establishment, we identified 22 protein-protein interactions and corresponding minimal interacting regions. A large fraction (32%) of the interactions we identified are supported by the literature. We did notice that, compared to the literature supported interactions, many novel interactions were identified in relatively few independent yeast colonies. One interpretation of this is that a more stringent cutoff could be used to increase the fraction of biologically relevant interactions. For example, for this particular dataset, keeping only those interactions that were found in 5 or more independent yeast colonies would raise the number of literature supported interactions from 32% to 50%. It is also possible, however, that interactions found in fewer independent yeast colonies represent interactions that are more difficult to detect, for example because the interaction is transient or weak. Literature derived interactions may have an inherent bias for more easily detected interactions. Removing the 10 interactions found fewer than 5 times would have eliminated one interaction supported by the literature, and three interactions validated by co-affinity purification. For this reason,

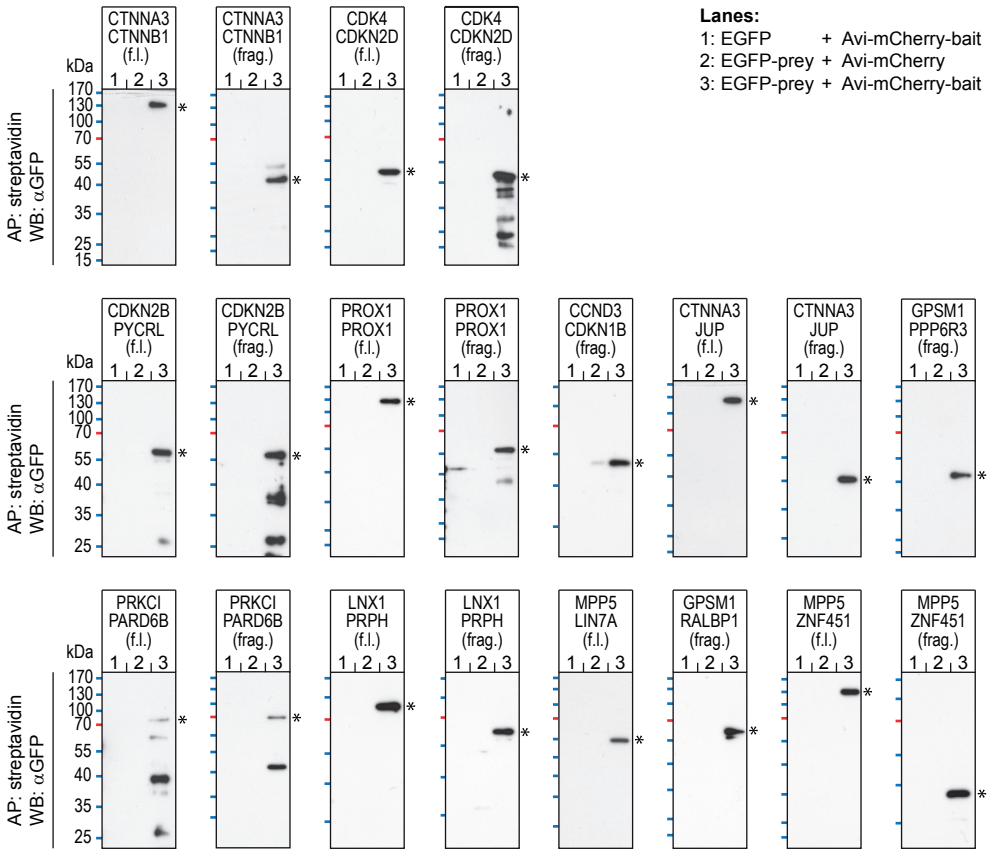


Figure 5. Interactions testing positive by co-affinity purification. Bait proteins are tagged with Avi-mCherry and prey proteins with EGFP. Depicted are αGFP Western blots on proteins purified using streptavidin coated beads. The tested proteins are indicated above each blot, the first protein listed is the bait, the second protein is the prey. Also indicated is whether the prey protein tested was full-length (f.l.) or corresponds to the shortest fragment identified by Y2H (frag.). In each blot, lanes 1 and 2 contain the negative control purifications, while lane 3 contains the actual co-affinity purification (1: Avi-mCherry-bait tested against the EGFP tag, 2: the EGFP-prey tested against the Avi-mCherry tag, 3: the interacting Avi-mCherry-bait and EGFP-prey). Asterisks indicate bands of the expected molecular mass. In some lanes, degradation products are visible as bands of lower size. Expression of tagged protein in input lysates is shown in Figure S2.

and because there is no experimental rationale for choosing a particular cutoff, we present all interactions, with the caveat that the interactions identified more frequently may be more easily reproduced in other assays.

In addition to the literature supported known interactions, 40% of the novel interactions are supported by independent co-affinity purification assays. The validity of the minimal interacting regions we identify is supported by the finding that we recapitulated the known interaction site for each of the literature described interactions. Moreover, 11 of the 12 without increasing the workload compared to traditional cDNA library screens. Initial library generation is the most time-consuming step. However, except for access to an

Table 3. Validation of interactions by co-affinity purification.

Bait	Prey			Interaction	
Name*	Name	Full-length size (a.a.)*	Fragment region (a.a.)*	Full-length	Fragment
CTNNA3 (130 kDa)	CTNNB1	781 (114 kDa)	64–165 (40 kDa)	yes	yes
CDK4 (64 kDa)	CDKN2D	166 (46 kDa)	1–142 (43 kDa)	yes	yes
CDKN2B (45 kDa)	PYCRL	286 (57 kDa)	20–285 (56 kDa)	yes	yes
PROX1 (113 kDa)	PROX1	737 (111 kDa)	184–381 (51 kDa)	yes	yes
DVL3 (108 kDa)	HOMEZ	550 (87 kDa)	52–550 (84 kDa)	no	no
CTNNA3 (130 kDa)	EHMT2	1210 (140 kDa)	907–1210 (63 kDa)	no	no
CCND3 (63 kDa)	CDKN1B	198 (50 kDa)	7–173 (47 kDa)	ND	yes
CDK4 (64 kDa)	CCND1	295 (62 kDa)	44–169 (43 kDa)	no	no
NUMA1 (139 kDa)	CCDC57	916 (114 kDa)	NA	no	NA
CTNNA3 (130 kDa)	JUP	745 (110 kDa)	53–173 (41 kDa)	yes	yes
GPSM1 (78 kDa)	PPP6R3	873 (126 kDa)	522–628 (40 kDa)	ND	yes
PRKCI (98 kDa)	PARD6B	372 (69 kDa)	10–372 (68 kDa)	yes	yes
CDK4 (64 kDa)	KLHL32	620 (82 kDa)	11–206 (50 kDa)	no	no
GPSM1 (78 kDa)	TRIM23	574 (92 kDa)	146–262 (41 kDa)	no	no
LNX1 (100 kDa)	PRPH	470 (82 kDa)	40–470 (78 kDa)	yes	yes
MPP5 (108 kDa)	LIN7A	233 (54 kDa)	22–100 (38 kDa)	yes	no
CDKN2B (45 kDa)	RNF20	975 (142 kDa)	232–466 (56 kDa)	no	no
CTNNA3 (130 kDa)	SPRY2	315 (67 kDa)	39–309 (58 kDa)	no	no
GPSM1 (78 kDa)	CBS	551 (89 kDa)	371–551 (49 kDa)	no	no
GPSM1 (78 kDa)	RALBP1	655 (104 kDa)	403–634 (56 kDa)	no	yes
MPP5 (108 kDa)	ZNF451	1061 (135 kDa)	306–395 (39 kDa)	yes	yes
PRKCI (98 kDa)	CRX	299 (61 kDa)	13–272 (56 kDa)	no	no

ND: Experiment could not be performed for technical reasons.

NA: Not applicable, no fragment smaller than full-length was identified by Y2H.

* molecular mass in kDa includes the EGFP or Avi-mCherry tag. Sizes and region coordinates are in amino acids.

ultrasonicator, the procedure uses basic molecular biology techniques and enzymes and can easily be adopted by others to generate random fragment libraries for any organism for which a source of cloned ORFs is available. The cloning of bait proteins and the screening procedure itself do not take any additional time compared with other library screens. Thus, our approach offers significant advantages without increasing the screening workload, apart from the initial investment in library generation.

MATERIALS AND METHODS

Y2H Gal4-AD vector

Vector pMB26 is a modified version of pPC86 that contains a flexible linker (GGSSGA) between the Gal4-AD and the cloned ORF, and facilitates blunt-end cloning using *Sma*I.[50] Gal4-AD is fused to the N-terminus of the ORF. pMB26 was generated by inserting an oligonucleotide linker into pPC86 digested with *Sal*I and NotI. The linker was created by annealing oligos pMB26_

F1: 5'- TCGAGTGGCGCGCCCGGGTAGC and pMB26_R1: 5'- GGCCGCTACCCGGGCGCGCCAC. pMB26 also contains the CYH2 gene from Clontech vector pAS2-1 (http://www.clontech.com). An EcoRV fragment containing CYH2 was cut from pAS2-1, and ligated into pPC86 digested with Acc65I and treated with Klenow + dNTPs to generate blunt ends.

Amplification of ORFeome clones
Each of the 15,483 ORF clones contained in the ORFeome 5.1 collection was PCR amplified using universal primers that anneal to the pDonr223 Gateway vector backbone (pDonr223_F: 5'-CCCAGTCACGACGTTGTAAAACG and pDonr223_R: 5'-GTAACATCAGAGATTTTGAGACAC). All PCR reactions were done in 384-well plates in a 15 µl volume, using Novagen KOD Hot Start DNA Polymerase. The reactions were assembled and run according to the manufacturer's instructions, using an annealing temperature of 58°C and an extension temperature of 68°C. The PCR products were collected in 5 pools based on size (<500 bp: 2610 ORFs, 500 – 1000 bp: 4811 ORFs, 1000 – 1500 bp: 3784 ORFs, 1500 – 2000 bp: 2116 ORFs, >2000 bp: 2162 ORFs), and purified using a Gel and PCR Clean-up kit (Machery-Nagel).

Removal of vector sequences
The PCR products contain ~75 bp of vector sequence preceding and following the ORF. These were removed by an Exonuclease III (Fermentas) treatment titrated to remove approximately 75 bp from each end. Each of the 5 PCR pools was incubated with 15 units Exonuclease III / pmol PCR product at 37°C for 8 minutes, and the reaction was stopped by heat inactivation at 75°C for 15 minutes. A 1 hour treatment with mung bean nuclease (1 Unit / µg DNA) (NEB) was used to remove the single strand overhangs left by the Exonuclease III treatment. The Exonuclease treated PCR pools were not further purified before proceeding with the fragmentation.

Fragmentation of PCR pools
To fragment PCR pools to the desired size range, 10 µg of each Exonuclease III treated pool of PCR products was sheared using a Covaris S2 Focused-ultrasonicator in a 130 µl microTUBE. Settings for 700 bp fragments: Intensity = 3, Duty Cycle = 5%, Cycles per Burst = 200, Treatment Time = 60 seconds. Settings for 500 bp fragments: Intensity = 3, Duty Cycle = 5%, Cycles per Burst = 200, Treatment Time = 80 seconds. Settings for 250 bp fragments: Intensity = 4, Duty Cycle = 10%, Cycles per Burst = 200, Treatment Time = 130 seconds. Each pool of fragments was then subjected to electrophoresis on an agarose gel, and DNA in the desired size range was isolated and purified using a Gel and PCR Clean-up kit (Machery-Nagel).

Ligation into pMB26
Fragmented DNA from each PCR pool was ligated into pMB26. To add a full-length component to the library, we also ligated into pMB26 DNA that had only been exonuclease treated, but not fragmented. The ends of the DNA to be ligated were converted to 5'-phosphorylated blunt ends using the Epicentre End-It DNA End-Repair Kit. Insert and vector were mixed in a 5:1 molar ratio, starting with 2 µg of insert DNA, in a final volume of 200 µl 1X T4 DNA ligase buffer containing 10 µl T4 ligase (50U) (Fermentas). For optimal ligation of blunt ends, the ligation reaction was cycled between 10°C and 30°C for 10 seconds each, for a total of 12 hours.

Transformation of ligated fragments into bacteria
Ligation mixtures were transformed into *E. coli* strain DH5α, made competent using the method of Inoue *et al*.[51] 100 µl of ligation mixture was mixed with 2 ml of competent bacterial cells, and split into 50 µl aliquots in a 96-well plate. The plate was incubated on ice for 30 minutes, heat shocked at 42°C for 45 seconds in a thermocycler, and placed back on ice for 2 minutes

before adding 150 µl of SOC medium to each well. All bacterial suspensions were collected and pooled in a 50 ml Erlenmeyer flask, and allowed to recover in a shaking incubator at 37°C for 1 hour. Finally, the transformations were plated on 15 cm ø LB-Agar plates containing 50 µg/ml Ampicillin (25 plates, 320 µl cells each), and grown overnight at 37°C.

For each pool, we determined the desired number of colonies to obtain. First, we calculated the number of fragments per ORF needed such that there is a 99% probability that every part of the full-length ORF is covered. The number of fragments needed is given by the expression N=ln(1-P)/ln(1-f), where N is the number of colonies, P is the probability, and f is the fraction of the full-length ORF covered by the fragment.[52] Thus, for the first pool (ORFs < 500 bp) aiming for a fragment size of 250 bp, N=ln(1-0.99)/ln(1-250⁄500), or 6.64 fragments/ORF. To obtain the number of colonies desired following ligation, the number of fragments/ORF is multiplied with the number of ORFs present in the pool, and multiplied by six to account for only 1/6 clones being in frame. For the full-length pools, we aimed for 10x the number of clones in the pool. When necessary, we repeated each ligation and transformation until sufficient colonies were reached.

Preparation of fragment library DNA

After O/N growth, bacterial colonies were washed off each plate in 5 ml LB medium containing 50 µg/ml Ampicillin, using a sterile cell scraper. Colonies from the same ligation reaction were all combined in a single Erlenmeyer flask, and LB medium containing 50 µg/ml Ampicillin was added to a final volume of 100 ml. After growing the culture for 2 hours at 37°C in a shaking incubator, plasmids were purified using a Maxi-prep kit (Machery-Nagel). DNA from each of the 5 fragmented PCR pools cloned into pMB26 was mixed such that all ORFs are represented at similar levels. DNA from each of the 5 full-length pools was mixed in a similar fashion. Full-length and fragment components were then mixed in a 1:3 ratio to obtain the final library.

Generating the Gal4-AD yeast mating library

To generate AD mating libraries for screening, yeast strain Y8800 (genotype *MAT***a** *trp1-901 leu2-3,112 ura3-52 his3-200 gal4Δ gal80Δ cyh2^R GAL1::HIS3@LYS2 GAL2::ADE2 GAL7::LacZ@met2*) was transformed with 90 µg of the complete AD fragment library, using the LiAc method.[53] Yeast were plated on SC-Trp plates to select for transformants. After two days of growth at 30°C, all colonies were harvested in 200 ml YEPD medium containing 20% glycerol (w/v), and frozen in 2 ml aliquots at -80°C.

Generating the Gal4-DB bait strains

For each of the ORFs to be used as a Gal4-DB bait fusion, the corresponding Gateway Entry clone plasmid was isolated from the ORFeome 5.1 collection. The correct identity of the ORF was verified by sequencing. Each ORF was then transferred into the pDest-DB Y2H vector by Gateway recombinational cloning (Life Technologies). The bait constructs were transformed into yeast strain Y8930 (genotype *MAT*α *trp1-901 leu2-3,112 ura3-52 his3-200 gal4Δ gal80Δ cyh2^R GAL1::HIS3@LYS2 GAL2::ADE2 GAL7::LacZ@met2*) using the LiAc method.[53] To test for autoactivation of Y2H reporter genes in the absence of an interacting AD-Y fusion protein, each bait strain was plated on an SC plate lacking leucine and histidine. Strains able to grow are able to activate the HIS3 reporter, and were eliminated from the screening process.

Y2H screening

All Y2H screens were done using a mating-based approach as previously described.[27] Each Gal4-DB bait strain was grown overnight at 30°C in 5 ml of YEPD medium in a shaking incubator.

An equal amount of Gal4-DB cells and thawed Gal4-AD fragment library was then mixed in a 15 ml conical tube. For each screen, we used an amount of yeast that would yield an O.D. 600 value of 6 if suspended in 1 ml (e.g., if the O.D. 600 of the overnight culture is 3, then 2 ml of this culture was used). Mixed yeast cells were spun down at 650 g for 3 minutes, resuspended in 200 µl of sterile water, and plated on a 10cm ø YEPD media plate. The plates were incubated at 30°C for 4 hours, which results in a low (~1%) mating efficiency. This minimizes the chance that yeast cells undergo mitosis after mating, which could lead to the identification of interactions from multiple colonies derived from the same parent yeast. After the incubation, yeast cells were washed off the plate in 200 µl sterile water, and plated on a 15cm ø SC –Leu –Trp –His media plate.

After 4 days of growth at 30°C, colonies growing on the SC –Leu –Trp –His plates were picked into a 96-well plate with 25 µl of water, and plated on two fresh SC –Leu –Trp –His plates. Controls of known reporter activity strength were also added to these plates. After 3 days of growth at 30°C, yeast from one of the two plates was used to identify the identity of the interacting AD fusion by PCR and sequencing. The other plate was replica plated to four different assay plates: SC –Leu –Trp –His and SC –Leu –Trp –His + 2 mM 3AT to gauge the strength of activation of the *HIS3* reporter gene, SC –Leu –Trp –Ade to gauge the strength of activation of the *ADE2* reporter gene, and SC –Leu –His + 1 µg/ml cycloheximide to identify autoactivating yeast colonies that express the *HIS3* reporter gene even in the absence of an AD fusion protein.[21] Yeast colonies that express at least one reporter gene, and which do not show growth on the corresponding autoactivation plates, are considered positive.

PCR amplification and sequencing of AD clones

From each positive yeast colony a small amount of cells, roughly a sphere of 0.5 mm diameter, was resuspended in 20 µl lysis buffer (0.1 M $NaPO_4$ buffer pH 7.4, 2.5 mg/ml Zymolase 20T, Seikagaku Corporation) in a 96-well plate, using 200 µl pipet tips. Plates with resuspended yeast were then incubated for 5 minutes at 37°C followed by 5 minutes at 94°C. Next, 80 µl of sterile water was added. From the lysed yeast cells, the ORF insert in the Gal4-AD plasmid was amplified using primers AD: 5'-CGCGTTTGGAATCACTACAGGG and TERM: 5'-GGAGACTTGACCAAACCTCTGGCG. The PCR reactions were done using Novagen KOD Hot Start DNA Polymerase, using 2 µl of lysed yeast in a 25 µl reaction. The reactions were assembled and run according to the manufacturer's instructions, using an annealing temperature of 58°C, an extension temperature of 68°C, and an extension time of 5 minutes. The amplified ORF inserts were then sent for sequencing using primer AD. All sequencing was done by Macrogen, in 96-well plate format with purification of PCR product performed by Macrogen.

Sequence data analysis

Each sequence read was analyzed as follows: first, phred was used to find the high quality segment of the read, using the -trim_alt option.[54] Next, the vector sequences were clipped from the 5'- and 3'- ends of the read, leaving only the insert sequence. Traces where we were unable to identify an exact match to the 12 bases of vector sequence directly preceding the insert were eliminated. Traces where the Gal4-AD sequence was not in frame with the insert sequence were also eliminated. The remaining traces were used to determine the identity of the interacting protein, by comparing the trace sequence to the ORF predictions in the human genome reference consortium release GRCh37_61 (obtained from the ensemble ftp site: ftp://ftp.ensembl.org) by BLAST+.[55] The start and endpoints of each insert relative to the corresponding full-length ORF

were determined by direct comparison of the sequence read with the predicted ORF sequence, or using phrap when an exact match could not be found (e.g., when the sequence trace is not fully accurate).[56]

For comparison of randomly sequenced fragments from the library with the full-length ORF, when multiple splice variants are predicted to exist, we plotted the fragment against the first splice variant in the database (alphabetically) that contains the entire fragment.

Mammalian expression constructs

Vectors used were pCI-NEO-BirA, Avi-mCherry-C1, and pEGFP-C1 (Clontech). Avi-mCherry-C1 contains the sequence MASGLNDIFEAQKIEWHEGGG, which is a substrate for the biotin ligase BirA, upstream of mCherry.[57,58] All ORFs were amplified from ORFeome clones by PCR, and cloned into Avi-mCherry-C1 or pEGFP-C1 using *Eco*RI/*Sal*I, *Eco*RI/*Kpn*I, *Sal*I/*Kpn*I, or *Sal*I/*Bam*HI sites added to the PCR primers.

Co-affinity purification

Constructs were transfected into HEK293 cells grown in DMEM/Ham's F10 (50/50%) medium containing 10% FCS and 1% penicillin/streptomycin. One day before transfection, nearly confluent cells were plated at 1:10 in 10 cm tissue culture dishes. Cells were transfected using polyethylenimine (PEI) as follows: plasmid DNA was diluted in 500 µl Ham's F10 medium, and PEI was added in a 3:1 PEI(µg):DNA(µg) ratio. After a 20 minute incubation at room temperature during which the cell culture medium was refreshed, the PEI/DNA mixture was added to the cells in a dropwise fashion. For every protein pair to be tested, 3 sets of plasmids were transfected. 1: Avi-mCherry-bait (7 µg) + empty EGFP vector (2 µg) + BirA (5 µg), 2: empty Avi-mCherry vector (2 µg) + EGFP-prey (7 µg) + BirA (5 µg), 3: Avi-mCherry-bait (7 µg) + EGFP-prey (7 µg) + BirA (5 µg).

Cells were harvested 24 hours after transfection, by washing and then scraping the cells in 2 ml ice-cold TBS (20 mM Tris, 150 mM NaCl, pH 8.0). Cells pelleted by centrifugation at 1000 g for 3 minutes were lysed in 500 µl lysis buffer (20 mM Tris-HCl, pH 8.0, 150 mM NaCl, 1.0% Triton X-100, and protease inhibitors; Roche), by freezing the cells at -80°C for 30 minutes, and incubating them for 30 minutes on ice after thawing at room temperature. Cell debris was removed by centrifugation at 13,200 rpm for 15 minutes. A 50 µl sample of cell lysate was used to check for expression of EGFP and mCherry constructs. The remainder was mixed with 25 µl of Dynabeads M-270 streptavidin (Invitrogen) first blocked by a 45 minute incubation with 0.2% chicken egg white in lysis buffer, and incubated for 1.5 hours rotating at 4°C. Beads were separated by centrifugation at 9000 rpm for 30 seconds, and washed three times in wash buffer (20 mM Tris-HCl, pH 8.0, 150 mM NaCl, 0.5% Triton X-100, and protease inhibitors; Roche). Bound proteins were eluted with 2x SDS sample buffer.

All protein samples were separated on 10% acrylamide gels, and subjected to western blotting on polyvinylidene difluoride membrane (Immobilon-P; Millipore). Blots were blocked with 5% skim milk in PBST (7 mM Na2HPO4, 3 mM NaH2PO4, 140 mM NaCl, 5 mM Kcl, 0.05% Tween-20) for 1 hour at room temperature. For detection of EGFP, blots were incubated with rabbit polyclonal anti-GFP (Abcam ab6556, 1:1000) in PBST + 5% skim milk for 1 hour at room temperature, washed with PBST three times for 10 minutes at room temperature, incubated with anti-rabbit IgG antibody conjugated to horseradish peroxidase (Jackson Immuno Research 111035003, 1:10,000) for 45 minutes at room temperature, washed with PBST three times for 10 minutes at room temperature, and finally washed once with PBS at room temperature for 10

minutes. For detection of biotinylated Avi-mCherry, blots were washed in PBST twice for 10 minutes at room temperature, incubated with Streptavidin coupled to horseradish peroxidase (Pierce High Sensitivity Streptavidin HRP, 1:20,000) in PBST for 1.5 hours at room temperature, washed in PBST five times for 10 minutes at room temperature, and finally washed once with PBS at room temperature for 10 minutes. Blots were developed using enhanced chemiluminescent Western blotting substrate (BioRad).

Domain predictions

Computationally predicted protein domains, as well as regions of low complexity and coiled coils, were obtained from SMART 7 (http://smart.embl-heidelberg.de/) and from Pfam 26.0.[59,60]

ACKNOWLEDGEMENTS

We thank Ewart de Bruin and Edwin Cuppen for use of and help with the Covaris ultrasonicator, Marina Mikhaylova and Anna Akhmanova for reagents and protocols for co-affinity purification, and David Hill and Marc Vidal for sharing the human ORFeome collection. We also thank Sander van den Heuvel for helpful discussions. The Human Interactome HI-2012 dataset was generated by the Center for Cancer Systems Biology (CCSB) at the Dana-Farber Cancer Institute, with funding from The National Human Genome Research Institute (NHGRI) of NIH; The Ellison Foundation, Boston, MA; and The Dana-Farber Cancer Institute Strategic Initiative. This work was funded by a Horizon grant from the Netherlands Organization for Scientific Research (NWO)/Netherlands Genomics Initiative (NGI) to M.B.

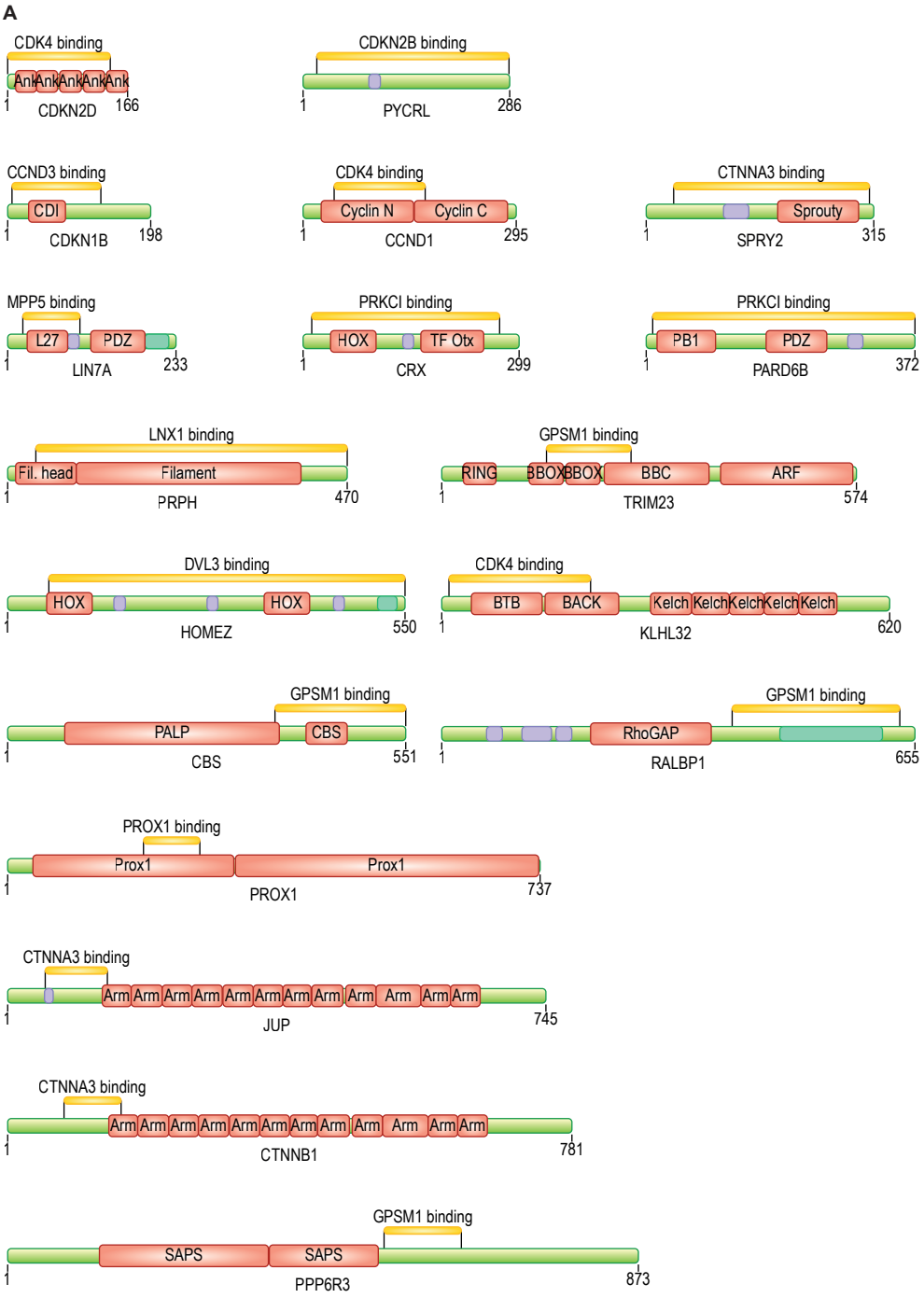
REFERENCES

1. Ponting, C. P. & Russell, R. R. The natural history of protein domains. Annu. Rev. Biophys. Biomol. Struct. 31, 45–71 (2002).
2. Bornberg-Bauer, E., Beaussart, F., Kummerfeld, S. K., Teichmann, S. A. & Weiner, J. The evolution of domain arrangements in proteins and interaction networks. Cell. Mol. Life Sci. 62, 435–45 (2005).
3. Jin, J. *et al*. Eukaryotic protein domains as functional units of cellular evolution. Sci. Signal. 2, ra76 (2009).
4. Pawson, T. & Nash, P. Assembly of cell regulatory systems through protein interaction domains. Science 300, 445–52 (2003).
5. Jadwin, J. A., Ogiue-Ikeda, M. & Machida, K. The application of modular protein domains in proteomics. FEBS Lett. 586, 2586–96 (2012).
6. Ho, Y. *et a*l. Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. 415, 2–5 (2002).
7. Gavin, A.-C. *et al*. Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature 415, 141–7 (2002).
8. Giot, L. *et al*. A protein interaction map of *Drosophila melanogaster*. Science 302, 1727–36 (2003).
9. Li, S. *et al*. A map of the interactome network of the metazoan *C. elegans*. Science 303, 540–3 (2004).
10. Rual, J.-F. *et al*. Towards a proteome-scale map of the human protein-protein interaction network. Nature 437, 1173–8 (2005).
11. Stelzl, U. *et al*. A human protein-protein interaction network: a resource for annotating the proteome. Cell 122, 957–68 (2005).
12. Yu, H. *et al*. High-quality binary protein interaction map of the yeast interactome network. Science 322, 104–10 (2008).
13. *Arabidopsis* Interactome Mapping Consortium. Evidence for network evolution in an *Arabidopsis* interactome map. Science 333, 601–7 (2011).

14. Luo, Q., Pagel, P., Vilne, B. & Frishman, D. DIMA 3.0: domain interaction map. Nucleic Acids Res. 39, D724–9 (2011).
15. Stein, A., Céol, A. & Aloy, P. 3Did: identification and classification of domain-based interactions of known three-dimensional structure. Nucleic Acids Res. 39, D718–23 (2011).
16. Yellaboina, S., Tasneem, A., Zaykin, D. V, Raghavachari, B. & Jothi, R. DOMINE: a comprehensive collection of known and predicted domain-domain interactions. Nucleic Acids Res. 39, D730–5 (2011).
17. Reich, S. *et al*. Combinatorial domain hunting: an effective approach for the identification of soluble protein domains adaptable to high-throughput applications. Protein Sci. 15, 2356–65 (2006).
18. Littler, E. Combinatorial domain hunting: solving problems in protein expression. Drug Discov. Today 15, 461–7 (2010).
19. Neveu, G. *et al*. Comparative analysis of virus-host interactomes with a mammalian high-throughput protein complementation assay based on Gaussia princeps luciferase. Methods 58, 349–59 (2012).
20. Rozenblatt-Rosen, O. *et al*. Interpreting cancer genomes using systematic host network perturbations by tumour virus proteins. Nature 487, 491–5 (2012).
21. Vidalain, P.-O., Boxem, M., Ge, H., Li, S. & Vidal, M. Increasing specificity in high-throughput yeast two-hybrid experiments. Methods 32, 363–70 (2004).
22. Braun, P. *et al*. An experimentally derived confidence score for binary protein-protein interactions. Nat. Methods 6, 91–7 (2009).
23. Venkatesan, K. *et al*. An empirical framework for binary interactome mapping. Nat. Methods 6, 83–90 (2009).
24. Chen, Y.-C., Rajagopala, S. V., Stellberger, T. & Uetz, P. Exhaustive benchmarking of the yeast two-hybrid system. Nat. Methods 7, 667–8 (2010).
25. Stellberger, T. *et al*. Improving the yeast two-hybrid system with permutated fusions proteins: the Varicella zoster virus interactome. Proteome Sci. 8, 8 (2010).
26. Worseck, J. M., Grossmann, A., Weimann, M., Hegele, A. & Stelzl, U. A stringent yeast two-hybrid matrix screening approach for protein–protein interaction discovery. Methods Mol. Biol. 812, 63–87 (2012).
27. Boxem, M. *et al*. A protein domain-based interactome network for *C. elegans* early embryogenesis. Cell 134, 534–45 (2008).
28. Lamesch, P. *et al*. hORFeome v3.1: a resource of human open reading frames representing over 10,000 human genes. Genomics 89, 307–15 (2007).
29. Degrado-Warren, J. *et al*. Construction and characterization of a normalized yeast two-hybrid library derived from a human protein-coding clone collection. Biotechniques 44, 265–73 (2008).
30. Rain, J.-C., Cribier, A., Gérard, A., Emiliani, S. & Benarous, R. Yeast two-hybrid detection of integrase-host factor interactions. Methods 47, 291–7 (2009).
31. Dricot, A. *et al*. Generation of the Brucella melitensis ORFeome version 1.1. Genome Res. 14, 2201–6 (2004).
32. Lamesch, P. *et al*. C. elegans ORFeome version 3.1: increasing the coverage of ORFeome resources with improved gene predictions. Genome Res. 14, 2064–9 (2004).
33. Brandner, C. J. *et al*. The ORFeome of Staphylococcus aureus v 1.1. BMC Genomics 9, 321 (2008).
34. Pellet, J. *et al*. ViralORFeome: an integrated database to generate a versatile collection of viral ORFs. Nucleic Acids Res. 38, D371–8 (2010).
35. Rajagopala, S. V *et al*. The *Escherichia coli* K-12 ORFeome: a resource for comparative molecular microbiology. BMC Genomics 11, 470 (2010).
36. Yang, X. *et al*. A public genome-scale lentiviral expression library of human ORFs. Nat. Methods 8, 659–61 (2011).
37. Kerrien, S. *et al*. The IntAct molecular interaction database in 2012. Nucleic Acids Res. 40, D841–6 (2012).
38. Aberle, H. *et al*. Single amino acid substitutions in proteins of the armadillo gene family abolish their binding to α-catenin. J. Biol. Chem. 271, 1520–6 (1996).
39. Hirano, Y. *et al*. Structure of a cell polarity regulator, a complex between atypical PKC and Par6 PB1 domains. J. Biol. Chem. 280, 9653–61 (2005).

40. Jeffrey, P. *et al*. Mechanism of CDK activation revealed by the structure of a cyclinA-CDK2 complex. Nature 376, 313–20 (1995).
41. Russo, A., Jeffrey, P., Patten, A., Massagué, J. & Pavletich, N. Crystal structure of the p27Kip1 cyclin-dependent-kinase inhibitor bound to the cyclin A-Cdk2 complex. Nature 382, 325–31 (1996).
42. Russo, A. A., Tong, L., Lee, J., Jeffrey, P. D. & Pavletich, N. P. Structural basis for inhibition of the the tumour suppressor p16 INK4a. Nature 395, 237–43 (1998).
43. Zhang, J. *et al*. Structure of an L27 domain heterotrimer from cell polarity complex Patj/Pals1/Mals2 reveals mutually independent L27 domain assembly mode. J. Biol. Chem. 287, 11132–40 (2012).
44. Massiah, M. A., Simmons, B. N., Short, K. M. & Cox, T. C. Solution structure of the RBCC/TRIM B-box1 domain of human MID1: B-box with a RING. J. Mol. Biol. 358, 532–45 (2006).
45. Trockenbacher, A. *et al*. MID1, mutated in Opitz syndrome, encodes an ubiquitin ligase that targets phosphatase 2A for degradation. Nat. Genet. 29, 287–94 (2001).
46. Zhong, W. *et al*. Lanthionine synthetase C-like protein 1 interacts with and inhibits cystathionine β-synthase: a target for neuronal antioxidant defense. J. Biol. Chem. 287, 34189–201 (2012).
47. Ahmad, K. F. *et al*. Mechanism of SMRT corepressor recruitment by the BCL6 BTB domain. Mol. Cell 12, 1551–64 (2003).
48. Gamsjaeger, R., Liew, C. K., Loughlin, F. E., Crossley, M. & Mackay, J. P. Sticky fingers: zinc-fingers as protein-recognition motifs. Trends Biochem. Sci. 32, 63–70 (2007).
49. Schapira, M. Structural Chemistry of Human SET Domain Protein Methyltransferases. Curr. Chem. Genomics 5, 85–94 (2011).
50. Chevray, P. M. & Nathans, D. Protein interaction cloning in yeast: identification of mammalian proteins that react with the leucine zipper of Jun. Proc. Natl. Acad. Sci. U. S. A. 89, 5789–93 (1992).
51. Inoue, H., Nojima, H. & Okayama, H. High efficiency transformation of *Escherichia coli* with plasmids. Gene 96, 23–28 (1990).
52. Clarke, L. & Carbon, J. A colony bank containing synthetic Col El hybrid plasmids representative of the entire *E. coli* genome. Cell 9, 91–9 (1976).
53. Schiestl, R. & Gietz, R. High efficiency transformation of intact yeast cells using single stranded nucleic acids as a carrier. Curr. Genet. 16, 339–46 (1989).
54. Ewing, B., Hillier, L., Wendl, M. C. & Green, P. Base-calling of automated sequencer traces using Phred. I. Accuracy assessment. Genome Res. 8, 175–85 (1998).
55. Camacho, C. *et al*. BLAST+: architecture and applications. BMC Bioinformatics 10, 421 (2009).
56. Rieder, M. J., Taylor, S. L., Tobe, V. O. & Nickerson, D. A. Automating the identification of DNA variations using quality-based fluorescence re-sequencing: analysis of the human mitochondrial genome. Nucleic Acids Res. 26, 967–73 (1998).
57. Lansbergen, G. *et al*. CLASPs attach microtubule plus ends to the cell cortex through a complex with LL5beta. Dev. Cell 11, 21–32 (2006).
58. Van Spronsen, M. *et al*. TRAK/Milton motor-adaptor proteins steer mitochondrial trafficking to axons and dendrites. Neuron 77, 485–502 (2013).
59. Letunic, I., Doerks, T. & Bork, P. SMART 7: recent updates to the protein domain annotation resource. Nucleic Acids Res. 40, D302–5 (2012).
60. Punta, M. *et al*. The Pfam protein families database. Nucleic Acids Res. 40, D290–301 (2012).

SUPPORTING INFORMATION



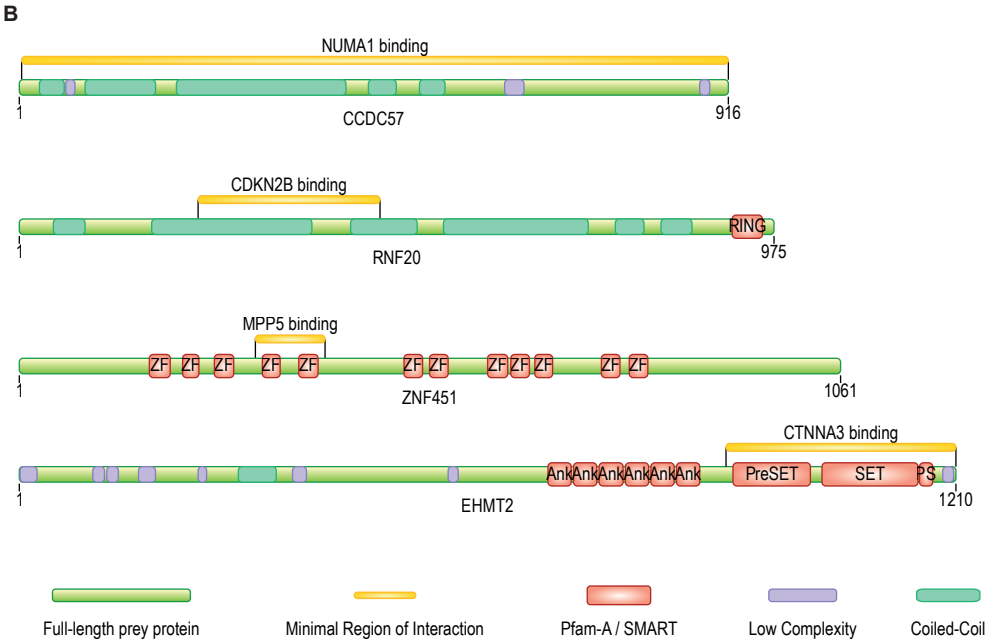
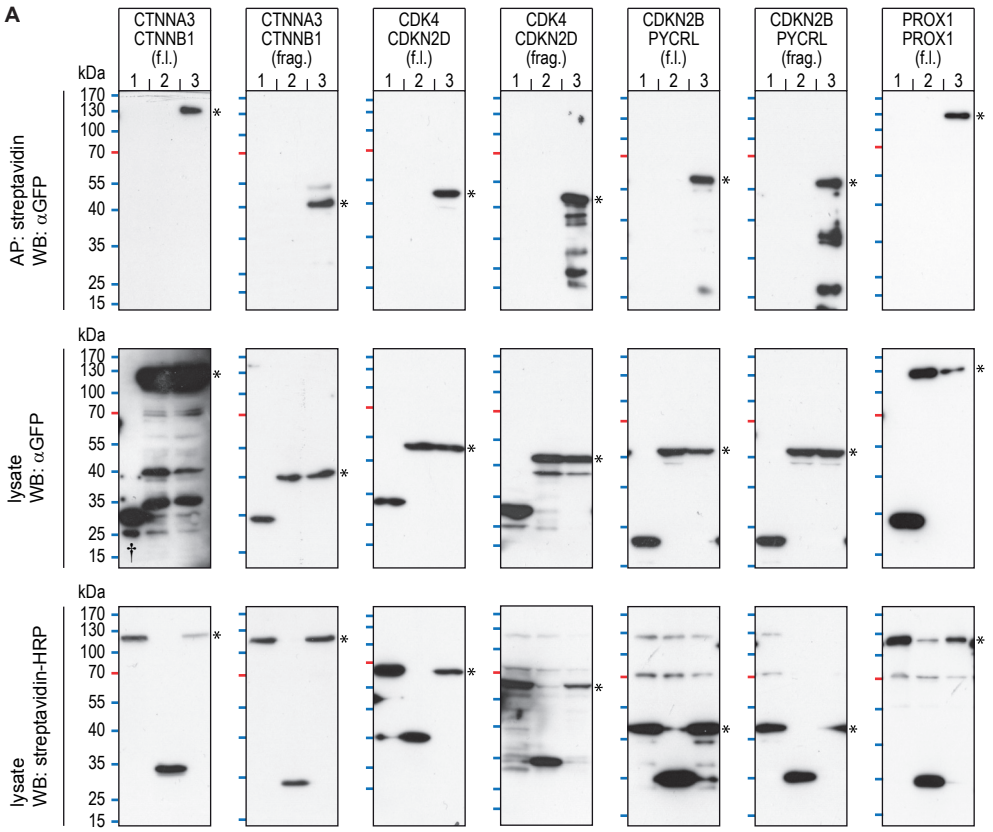
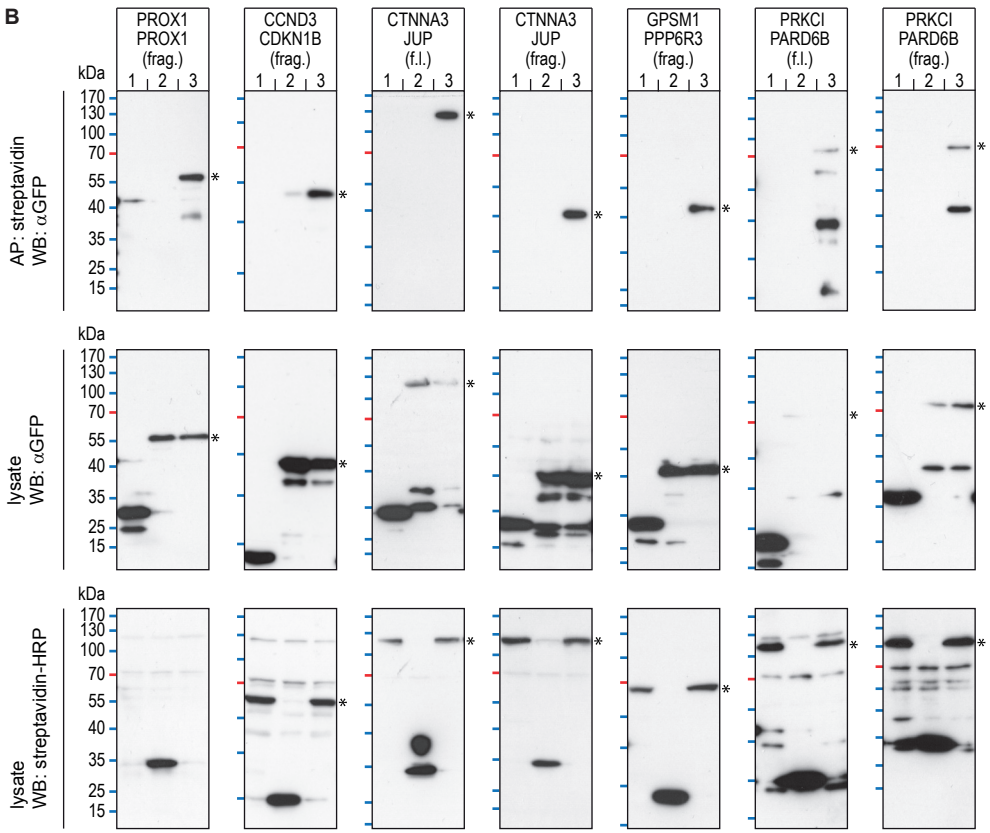


Figure S1. Graphical representation of minimal region of interaction for all interacting protein pairs. (A and B) Green bars represent full-length proteins. Yellow bars represent regions of the full-length protein required for interaction with the indicated binding partner. Pfam-A and SMART domain signatures are drawn as red boxes, coiled-coil predictions as green boxes, and low complexity regions as blue boxes. Where multiple predictions overlapped, display preference for these domains was in the following order: Pfam-A/SMART, coiled-coil, low complexity.

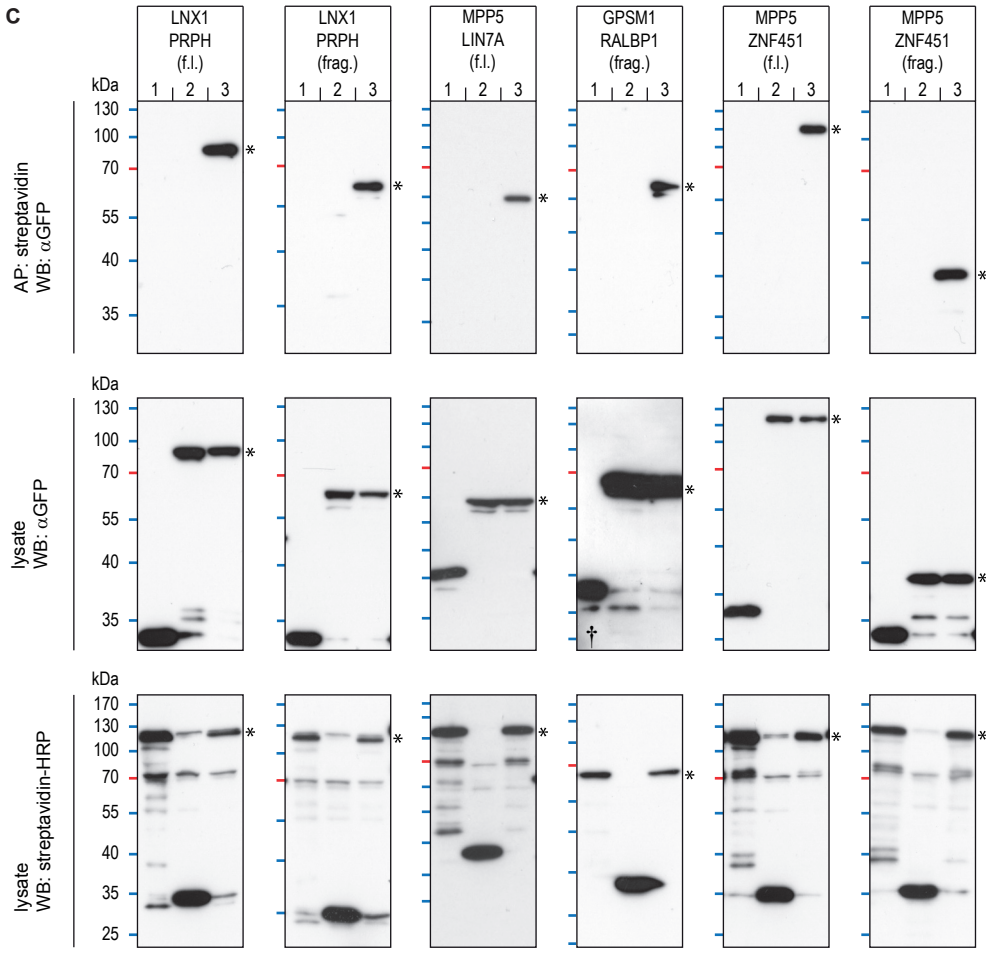


Lanes:
1: EGFP + Avi-mCherry-bait
2: EGFP-prey + Avi-mCherry
3: EGFP-prey + Avi-mCherry-bait

Figure S2. Interactions testing positive by co-affinity purification. (A, B, and C) Bait proteins are tagged with Avi-mCherry and prey proteins with EGFP. For each protein pair, 3 blots are shown. Top: αGFP Western blots on proteins purified using streptavidin coated beads, middle: αGFP Western blots on input lysates (to determine EGFP expression levels), bottom: streptavidin-HRP Western blots on input lysates (to determine expression and biotinylation of Avi-mCherry). The tested proteins are indicated above each series of blots, the first protein listed is the bait, the second protein is the prey. Also indicated is whether the prey protein tested was full-length (f.l.) or corresponds to the shortest fragment identified by Y2H (frag.). In each blot, negative control mixtures are in lanes 1 and 2, while lane 3 contains the bait and prey proteins that interacted in the actual co-affinity purification (1: Avi-mCherry-bait tested against the EGFP tag, 2: the EGFP-prey tested against the Avi-mCherry tag, 3: the interacting Avi-mCherry-bait and EGFP-prey). Asterisks indicate EGFP-prey and Avi-mCherry-bait bands of the expected molecular mass. Empty Avi-mCherry (lane 2 in bottom blots) and EGFP (lane 1 in middle blots) both have a mass of 29 kDa, and are not indicated. In some lanes, degradation products are visible as bands of lower size. In the streptavidine-HRP Westerns (bottom row of blots), endogenously biotinylated proteins result in the detection of background bands in addition to the biotinylated Avi-mCherry fusion proteins. In two experiments, expression of the EGFP-prey fusion proteins was low, and 1/5 of empty-EGFP input lysate was loaded to prevent overexposure in the overnight exposure used to detect EGFP-prey fusion levels (indicated with a † in the blot). Finally, in the PROX-1 – PROX-1(frag) transfections, levels of biotinylated Avi-mCherry::PROX-1 were below the limits of detection in input lysate, but detectable by immunofluorescence microscopy and sufficient for affinity purification of EGFP::PROX-1.



Lanes:
1: EGFP + Avi-mCherry-bait
2: EGFP-prey + Avi-mCherry
3: EGFP-prey + Avi-mCherry-bait



Lanes:
1: EGFP + Avi-mCherry-bait
2: EGFP-prey + Avi-mCherry
3: EGFP-prey + Avi-mCherry-bait

Table S1. Interactions identified in our screens for which interaction domain information is present in the literature.

Bait name	Bait description	Prey name	Prey description	Link to publication detailing interaction domain
CTNNA3	Alpha Catenin	CTNNB1	Beta Catenin	http://www.ncbi.nlm.nih.gov/pubmed/8576147
CDK4	Cdk4	CDKN2D	p19Ink4d	http://www.ncbi.nlm.nih.gov/pubmed/9751050
CCND3	Cyclin D3	CDKN1B	p27Kip1	http://www.ncbi.nlm.nih.gov/pubmed/8684460
CDK4	Cdk4	CCND1	Cyclin D1	http://www.ncbi.nlm.nih.gov/pubmed/7630397
CTNNA3	Alpha Catenin	JUP	Junction Plakoglobin	http://www.ncbi.nlm.nih.gov/pubmed/8576147
PRKCI	Protein kinase C, iota	PARD6B	Par6	http://www.ncbi.nlm.nih.gov/pubmed/15590654
MPP5	Pals1	LIN7A	LIN7	http://www.ncbi.nlm.nih.gov/pubmed/22337881

Table S2. Minimal interacting regions identified. All coordinates are in basepairs. Corresponding images are in Figure S1.

Bait	Prey	Prey splice	MRI start	MRI end	Prey ORF length
CDK4	CDKN2D	CDKN2D-201	1	142	166
CCND3	CDKN1B	CDKN1B-001	7	129	198
MPP5	LIN7A	LIN7A-201	22	100	233
CDKN2B	PYCRL	PYCRL-001	20	285	286
CDK4	CCND1	CCND1-201	44	169	295
PRKCI	CRX	CRX-201	13	272	299
CTNNA3	SPRY2	SPRY2-001	39	309	315
PARD3	CREB3	CREB3-001	74	271	371
PRKCI	PARD6B	PARD6B-001	10	372	372
LNX1	PRPH	PRPH-001	40	470	470
DVL3	HOMEZ	HOMEZ-201	58	550	550
GPSM1	CBS	CBS-001	371	551	551
GPSM1	TRIM23	TRIM23-001	146	262	574
CDK4	KLHL32	KLHL32-001	11	206	620
GPSM1	RALBP1	RALBP1-001	403	634	655
PROX1	PROX1	PROX1-001	189	266	737
CTNNA3	JUP	JUP-001	53	138	745
CTNNA3	CTNNB1	CTNNB1-001	82	141	781
GPSM1	PPP6R3	PPP6R3-001	522	628	873
NUMA1	CCDC57	CCDC57-001	4	916	916
CDKN2B	RNF20	RNF20-001	232	466	975
MPP5	ZNF451	ZNF451-001	306	395	1061
CTNNA3	EHMT2	EHMT2-001	914	1210	1210

Chapter 3

CRISPR/Cas9-targeted mutagenesis in *Caenorhabditis elegans*

Selma Waaijers[1], Vincent Portegijs[1], Jana Kerver[1], Bennie B. C. G. Lemmens[2], Marcel Tijsterman[2], Sander van den Heuvel[1] & Mike Boxem[1]

Genetics (2013) Nov 1; 195(3): 1187-91.

[1] Department of Biology, Utrecht University, 3584 CH Utrecht, The Netherlands.
[2] Department of Toxicogenetics, Leiden University Medical Center, 2333 ZC Leiden, The Netherlands.

ABSTRACT

The generation of *Caenorhabditis elegans* mutants has long relied on the selection of mutations in large-scale screens. Targeting specific loci in the genome for mutagenesis would greatly speed up analysis of gene function. Here, we adapted the CRISPR/Cas9 system to generate mutations at specific sites in the *C. elegans* genome.

INTRODUCTION

Current methods to generate mutations in the genome of *Caenorhabditis elegans*, including chemical mutagenesis and imprecise excision of transposons, all rely on recovering mutations in large-scale mutagenesis screens. Recently, several groups reported the use of the *Streptococcus pyogenes* CRISPR/Cas9 system to generate double strand break (DSB) induced mutations at specific genomic loci in organisms including yeast, flies, mammalian cells, and zebrafish.[1–7] Because of the enormous potential for targeted genome engineering, we here investigate the suitability of the CRISPR/Cas9 system for use in *C. elegans*. This article is one of six companion articles, that present different approaches to and features of CRISPR/Cas9 genome editing in *C. elegans*.[8–12]

The *S. pyogenes* CRISPR/Cas9 system effects site-specific cleavage of double stranded DNA through a complex containing the Cas9 endonuclease and two noncoding RNAs, crRNA and tracrRNA.[13,14] Target site specificity is mediated by a 20 nt region in the crRNA that is complementary to the target DNA, and a 3 nucleotide motif (NGG) following the target site in the DNA (termed PAM, for protospacer adjacent motif).[13,14] Thus a wide range of target sites can be chosen. Conveniently, a single synthetic guide RNA (sgRNA) that fuses the 3' end of crRNA to the 5' end of tracrRNA is sufficient to target Cas9 to a specific site and generate DSBs (Fig. 1A).[13]

RESULTS

To promote expression of Cas9, we codon-optimized the *S. pyogenes* Cas9 coding sequence for *C. elegans*, introduced artificial introns, and attached SV40 and *egl-13* nuclear localization signals to the N- and C-terminus, respectively, of the encoded Cas9 protein (Fig. 1B). To express Cas9 in the germline, we placed the Cas9 coding sequence under control of the *eft-3* or *hsp-16.48* promoters and the *tbb-2* 3'UTR, each of which has been shown to be compatible with germline expression.[15–17] To visualize expression of Cas9, we also generated Cas9::EGFP fusion vectors. We did not detect EGFP expression after injection of *Peft-3::Cas9::EGFP* (>20 animals examined). Injection of *Phsp-16.48::Cas9::EGFP* did result in visible EGFP expression, 5 hours after heat-shock induction for 1 hour at 34°C. Expression did vary between experiments: one series of injections resulted in high expression in 5/5 animals examined (Fig. 1C), while a second series of injections showed only weak expression in 1/12 animals examined. Because even low expression levels may provide sufficient Cas9 activity, we tested both *Peft-3* and *Phsp-16.48* containing constructs for activity in further experiments.

To provide the sgRNA, we tested two different approaches. First, we generated a vector containing a T7 promoter upstream of the sgRNA sequence for *in vitro* transcription of the sgRNA. Second, we generated a vector expressing the sgRNA under control of the regulatory sequences of an RNA polymerase III transcribed U6 snRNA on chromosome III, to enable *in vivo* transcription.[18] Both vectors contain *Bsa*I restriction sites for insertion of the target recognition sequence as an oligomer linker (Fig. 1D and E).

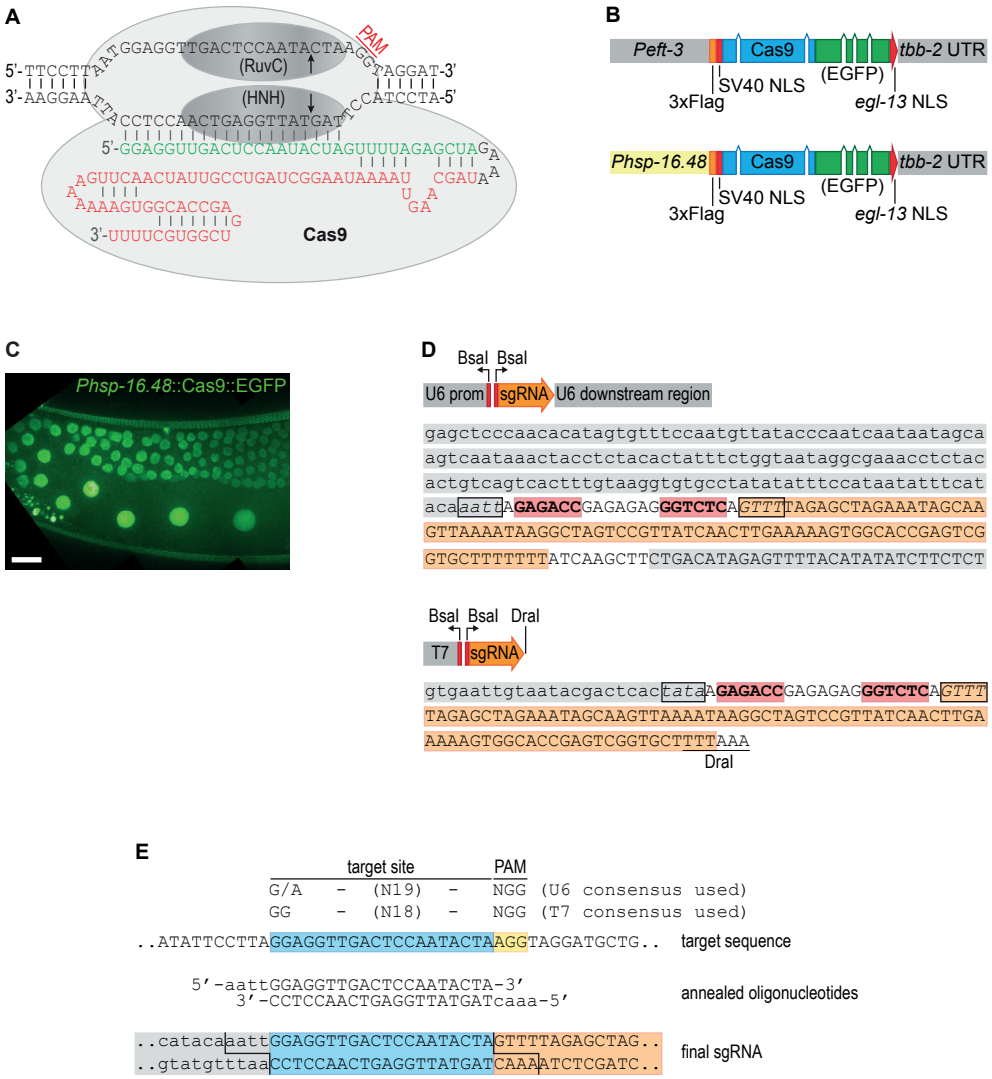


Figure 1. Experimental design and germline Cas9 expression. (A) Cas9/sgRNA in complex with a target site. RuvC and HNH endonuclease domains together generate a double strand break. In the sgRNA sequence, green bases are crRNA derived, and red bases tracrRNA derived. (B) Schematic of the Cas9 expression vectors used in this study, placing Cas9 or Cas9::EGFP under control of the *eft-3* promoter or the *hsp-16.48* heat shock promoter. Versions lacking EGFP are not shown. (C) Germline expression and nuclear localization of Cas9::EGFP expressed from the *hsp-16.48* heat shock promoter. Shown is a maximum intensity projection of a Z-stack. Scale bar 10 µm. (D) Diagrams and sequences of the U6::sgRNA and T7::sgRNA vectors. Grey background: promoter or downstream regions. Orange background: sgRNA sequence downstream of the target recognition sequence. Red background: *Bsa*I recognition sites. Boxed nucleotides: sequences left as 5' overhang after *Bsa*I digestion. (E) Example of cloning a target sequence into the U6::sgRNA vector. The 20bp target site is outlined in blue, and the PAM in yellow. The consensus sequences we used for target site selection are also indicated.

As a first test of functional activity, we generated a reporter construct carrying an out-of-frame copy of EGFP and *lacZ* downstream of the *myo-2* promoter. Imprecise repair of a DSB in a linker region between the first ATG and EGFP can result in a frameshift leading to EGFP expression. We co-injected the reporter (15 ng/µl) with *Peft-3::Cas9* or *Phsp-16.48::Cas9* (50 ng/µl), a U6 driven sgRNA targeting the linker region (50 ng/µl), and a *Pmyo-3::mCherry* co-injection marker (5 ng/µl). We also tested injection of lower Cas9/sgRNA concentrations (20 ng/µl both) together with *Pst*I digested λ DNA (20 ng/µl), to promote generation of more complex extrachromosomal arrays. Per condition we injected 10 animals, and *Phsp* expression was induced by a 1 hour heat shock at 34°C after injection. None of the injections with *Peft-3::Cas9* yielded viable transgenic F1. Instead, we observed mCherry expressing dead embryos, indicating a deleterious effect of this construct. A series of test injections showed that the embryonic lethality is concentration dependent, ranging from 30% at 1 ng/µl to 100% at 20 ng/µl (see Table S1). In contrast, 89% of the transgenic lines obtained from the injections with *Phsp-16.48::Cas9* expressed EGFP in the pharynx, indicating the presence of an extrachromosomal array with at least one frame-shifted copy of the reporter (Table 1). The injection of a lower concentration *Phsp::Cas9* diluted with lambda DNA resulted in a higher number of transgenic offspring, though the fraction of animals expressing EGFP was similar (90 and 84% resp., Table 1). Control injections lacking the sgRNA did not show EGFP expression, demonstrating specific Cas9/sgRNA activity (50 transgenic F1 examined).

We examined 18 stable transgenic lines obtained from EGFP expressing F1 animals. Of these, 15 expressed EGFP in most (>90%) of the F2 transgenic animals. The small fraction of EGFP negative transgenics could be due to mosaic inheritance of the extrachromosomal reporter array. Since Cas9 expression is induced by heat shock only in the injected P0 animals, these findings may indicate that DSBs were generated in the germline of the P0. Taken together, Cas9/sgRNA appears to efficiently generate DSBs in our plasmid-based reporter.

We next wanted to determine whether CRISPR/Cas9 can be used to generate heritable mutations at a specific genomic locus in *C. elegans*. For this purpose, we generated sgRNA constructs targeting the *lin-5* coding sequence near the known *ev571* mutation. We injected *Phsp::Cas9* together with either *in vitro* transcribed sgRNA or the U6::sgRNA plasmid, as well as the *Pmyo-3::mCherry* co-injection marker (Table 2). For each combination we injected 20 P0 animals, selected individual F1 animals expressing mCherry, and examined

Table 1. Number of transgenic and EGFP expressing F1 animals obtained using Cas9/sgRNA directed against an EGFP frameshift reporter.

			Results	
sgRNA conc.*	***Phsp-16.48::Cas9* conc.***	**No. of P0 injected**	**Transgenic F1**	**F1 expressing EGFP**
20	20	10	126	114 (90%)
50	50	10	32	27 (84%)

* All concentrations are in ng/µl. Injections with 20 ng/µl Cas9/sgRNA were supplemented with 20 ng/µl of *Pst*I digested λ DNA. All injections included 5 ng/µl of the *Pmyo-3::mCherry* marker to identify transgenic animals, and 15 ng/µl of the out-of-frame EGFP reporter.

Table 2. Number of transgenic F1 and mutant F2 progeny produced using Cas9/sgRNA directed against genomic loci.

sgRNA				Transgenic F1	
method/target	**conc.***	***Phsp-16.48::Cas9* conc.***	**No. of Po injected**	**No. selected**	**With mutant progeny**
U6 x *lin-5*	20	20	20	92	5
U6 x *lin-5*	50	50	20	24	5
T7 x *lin-5*	10	50	20	29	0
T7 x *lin-5*	150	50	20	124	0
U6 x *rol-1*	20	20	40	144	1
U6 x *rol-1*	50	50	20	140	2
U6 x *dpy-11*	50	50	20	20	2
U6 x *unc-119*	50	50	20	41	2

* All concentrations are in ng/µl. Injections with 20 ng/µl Cas9/sgRNA were supplemented with 35 ng/µl of *Pst*I digested λ DNA. All injections include 5 ng/µl of the *Pmyo-3::mCherry* marker to identify transgenic animals.

their F2 progeny for the presence of Lin-5 offspring. Animals injected with *in vitro* produced sgRNA failed to produce *lin-5* mutants (Table 2). In contrast, injections with *U6::sgRNA* yielded a total of 10 F1 animals that produced approximately ¼ Lin-5 offspring (Table 2). We confirmed the presence of mutations at the *lin-5* locus by sequence analysis, identifying several deletions and a 7 bp insertion (Fig. 2). For each F1 line we sequenced two mutant F2 animals independently, and in each case both animals harbored exactly the same mutation, strongly suggesting that the mutations were inherited from the parent, and were not generated *de novo* by somatic events. Two mutations could not be resolved: Sanger sequencing traces from both sides degrade into double peaks at the sgRNA target site. This can result from the presence of a repeated sequence, and we speculate that DSB repair resulted in the duplication of a short DNA sequence. Injections with the lower concentration of Cas9 and sgRNA expression plasmids coupled with lambda DNA yielded higher numbers of transgenic F1 animals, but ultimately produced the same number of *lin-5* mutants (Table 2).

Finally, we targeted three additional loci – *dpy-11*, *rol-1*, and *unc-119* – using *Phsp-16.48::Cas9* and *U6::sgRNA* (Table 2). As for *lin-5*, we selected transgenic F1 animals, and looked for the presence of visible mutants in the F2 generation. For *dpy-11* and *unc-119*, we identified two transgenic F1 each that segregated approximately ¼ mutant progeny, from a total of 20 and 41 transgenic F1 animals selected, respectively (Table 2). Homozygous mutations in *dpy-11* or *unc-119* were readily identified in all cases (Fig. 2). For *rol-1*, from 284 transgenic F1, we observed 3 plates with only a single Rol F2 animal. Sequencing of these mutants did confirm the presence of mutations at the target site (Fig. 2). It appears therefore that the *rol-1* phenotypes generated by our sgRNA are only partially penetrant. Together, these results confirm the ability of our approach to generate mutations at specific loci in the genome.

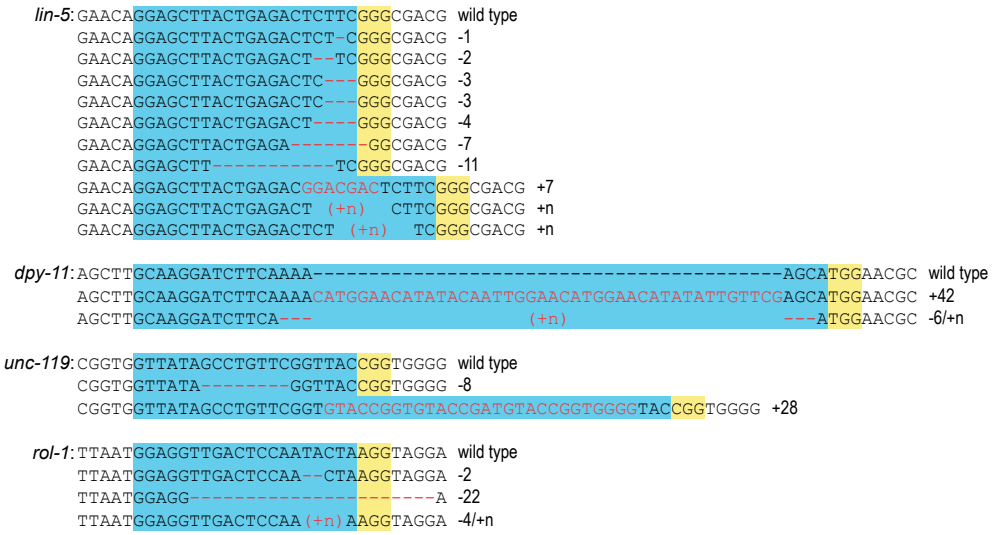


Figure 2. Genomic mutations generated by Cas9/sgRNA. Mutations are shown relative to the wild-type sequences. Three mutations could not be resolved by sequencing, and may correspond to insertion of a repeated sequence. Blue indicates sgRNA target site, and yellow is the PAM motif.

DISCUSSION

Here, we adapted the CRISPR/Cas9 system for use in *C. elegans*, and demonstrate its ability to efficiently generate genomic mutations. For *dpy-11*, *lin-5*, and *unc-119*, we obtained on average one mutant from every 5 or 6 P0 animals injected. For *rol-1*, the frequency was much lower (3 mutants out of 60 P0 injections), but the partial penetrance of the phenotype likely caused us to miss several mutations. The approach is not only efficient but also fast: cloning, mutant isolation, and sequencing of mutations can be completed in 10 days.

A recently published CRISPR/Cas9 method for *C. elegans* used *Peft-3* to drive Cas9 expression.[19] However, we found that expression of Cas9 from the *eft-3* promoter causes embryonic lethality. This contrasting result may be due to differences in the exact Cas9 protein produced. While the reason for the observed lethality is unclear, use of the heat shock promoter to provide a pulse of expression only in the injected animal circumvented this problem.

Five companion articles also report the successful application of CRISPR/Cas9 in *C. elegans*.[8–12] These groups used various approaches to provide Cas9 and sgRNA, including injection of Cas9 RNA or protein, and *in vitro* produced sgRNA. Thus, although in our case heat-shock induced Cas9 coupled with U6 driven sgRNA proved most efficient, it appears that the methodology to provide these two components can be highly flexible.

In addition to generating mutants, the DSBs produced by Cas9/sgRNA enable several other applications of genome engineering, including insertion of exogenous DNA through homologous recombination, and are likely to become an important tool for *C. elegans* researchers.[9,10]

MATERIALS AND METHODS

Culture conditions and strains

The wild-type *C. elegans* strain N2 was maintained under standard culture conditions as previously described.[20] All experiments were performed at 25°C, unless otherwise noted.

Plasmid construction

We generated four Cas9 expression constructs: pMB62 and pMB63 drive expression from the *eft-3* promoter, and are identical except for the presence of EGFP in pMB62. pMB66 and pMB67 drive expression from a heat shock promoter, and are again identical except for the EGFP fusion in pMB66. To generate these expression constructs, we first amplified the *tbb-2* 3' UTR from N2 genomic DNA by PCR, using a forward primer (5'-AAGAATTCATGCAAGATCCTTTCAAGCA), and reverse primer (5'-AAGAGCTCTGATCCACGATCTGGAAGATT) with EcoRI and SacI restriction sites, respectively. The resulting PCR product was cloned into pBluescript SK(+) digested with EcoRI and SacI. Next, we PCR amplified the *eft-3* promoter from pCFJ601 using a forward primer containing SalI (5'- AAGTCGACGCACCTTTGGTCTTTTATTGTCA), and a reverse primer containing XbaI and EcoRI sites (5'-AAGAATTCCCCGGGTCTAGATGAGCAAAGTGTTTCCCAACTG).[17] The resulting PCR product was cloned into the *tbb-2* construct digested with SalI and EcoRI. To add Cas9, we ordered a synthetic plasmid containing the 3xFlag tag, the SV40 NLS, the Cas9 coding sequences with artificial introns, and the *egl-13* NLS (Genscript). All sequences were codon optimized for *C. elegans* using the *C. elegans* Codon Adapter.[21] These sequences were cloned between the *eft-3* promoter and *tbb-2* 3' UTR using XbaI and EcoRI sites also present in the synthetic plasmid, resulting in vector pMB63. A unique SpeI site was added following the *tbb-2* UTR to facilitate future cloning efforts, by ligating a short oligonucleotide linker into an existing NdeI site. To generate pMB62, *C. elegans* optimized EGFP coding sequences were PCR amplified from pMA-mEGFP (a kind gift from Tony Hyman) using primers containing PstI sites (5'-AACTGCAGATGTCCAAGGGAGAGGAGCTC and 5'-AACTGCAGCTTGTAGAGCTCGTCCATTCCGTG), and cloned downstream of Cas9 using PstI. To generate the heat-shock expression constructs pMB66 and pMB67, the heat shock promoter *Phsp-16.48* was amplified from vector pJL44 using a forward primer containing a KpnI site (5'-AAGGTACCGCTGGACGGAAATAGTGGTAAAG) and a reverse primer containing an SpeI site (5'-AAACTAGTTCTTGAAGTTTAGAGAATGAACAGTAA), and inserted into pMB62 and pMB63 from which the *eft-3* promoter was removed using KpnI and XbaI (SpeI and XbaI digestions result in compatible overhangs).[15]

To generate the T7 sgRNA vector pMB60, the T7 promoter sequence followed by the BsaI cloning sites and the chimeric crRNA-tracrRNA sequences were ordered as a gBlocks Gene Fragment (IDT), and cloned blunt into cloning vector pMK digested with PvuII. To generate the U6 sgRNA vector pMB70, the U6 promoter sequence followed by the BsaI cloning sites and the chimeric crRNA-tracrRNA sequences were ordered as a gBlocks Gene Fragment (IDT), and cloned blunt into cloning vector pBluescript SK+ digested with EcoRV.[18] The sgRNA sequences were then transferred from pBluescript to pMK using PvuII sites present in both vectors. Finally, to add potential 3' regulatory sequences, we PCR amplified and inserted an 888 bp region downstream of the U6 snRNA using primers containing HindIII (5'- AAGCTTCTGACATAGAGTTTTACATATATCTTCTCTG) and SalI(5'- GTCGACCGAAGAGCACAGAAAAATTGG).

The Cas9 activity reporter plasmid pLM47 (*Pmyo-2::ATG::sgRNA target::EGFP::lacZ::unc-54UTR*) was constructed by replacing the C23 microsatellite of a previously generated microsatellite

instability reporter (pLM3, sequence available upon request) with an oligonucleotide linker containing a suitable sgRNA target sequence (GGATAACAGGGTAATTCTACCGG). The EGFP and LacZ coding sequences are out of frame with the first ATG, and require Cas9/sgRNA induced mutagenesis to be expressed.

sgRNA target site selection and cloning

The selection of a suitable sgRNA target site is limited by two requirements. First, the three nucleotides immediately following the target site have to correspond to the PAM consensus sequence of NGG (note that these three nucleotides are not actually incorporated in the sgRNA). Second, the promoters used may impose restrictions on the 5' nucleotides. In our case, efficient transcription from the T7 promoter is promoted by the incorporation of GG as the first two nucleotides of the RNA produced, while optimal transcription from a polymerase III promoter appears to require a purine as the first nucleotide of the RNA.[22–24] We therefore used the following sgRNA consensus sites: G/A-(N19)-NGG for the U6 vector, and GG-(N18)-NGG for the T7 vector. Though we chose to use these conservative consensus sites, it may be possible to ease the restrictions on the 5' nucleotides by using different promoters (especially for *in vivo* production of the sgRNA), or by extending the sgRNA sequence on the 5'-end with one or two nucleotides that do not participate in target recognition. To find suitable sites in the *dpy-11*, *lin-5*, *rol-1*, and *unc-119* genomic sequences, we searched for these consensus sequences using ApE – A plasmid Editor (http://biologylabs.utah.edu/jorgensen/wayned/ape/).

To facilitate cloning of different target sites into our vectors, we designed these to be digested with BsaI, a restriction enzyme that cuts outside of the recognition sequence. Two BsaI sites are juxtaposed such that upon digestion, the recognition sites themselves are eliminated, and two overhangs are created that exactly match the last four nucleotides of the U6 or T7 promoter, and the first four nucleotides of the sgRNA sequence. To insert the target sites, we ordered phosphorylated forward and reverse oligonucleotides that can be annealed to generate linkers compatible with BsaI digested T7 or U6 vector. For *lin-5*: lin-5_T7_Fwd: 5'-tataGGAGCTTACTGAGACTCTTC, lin-5_U6_Fwd: 5'-aattGGAGCTTACTGAGACTCTTC, and lin-5_Rev: 5'-aaacGAAGAGTCTCAGTAAGCTCC. For *rol-1*: rol-1_T7_Fwd: 5'-tataGGAGGTTGACTCCAATACTA, rol-1_U6_Fwd: 5'- aattGGAGGTTGACTCCAATACTA, and rol-1_Rev: 5'-aaacTAGTATTGGAGTCAACCTCC. For *dpy-11*: dpy-11_U6_Fwd: 5'-aattGCAAGGATCTTCAAAAAGCA and dpy-11_Rev: 5'-aaacTGCTTTTTGAAGATCCTTGC. For *unc-119*: unc-119_U6_Fwd: 5'-aattGTTATAGCCTGTTCGGTTAC and unc-119_Rev: 5'-aaacGTAACCGAACAGGCTATAAC. Oligonucleotides were annealed by heating 0.5 µmol of each oligonucleotide in annealing buffer (25 mM Tris-HCl, 10 mM $MgCl_2$, pH 7.5) and slowly cooling to room temperature. Annealed oligonucleotides were ligated in vectors digested with BsaI, and inserts were verified by sequencing.

***In vitro* transcription**

In vitro transcribed sgRNA was generated with the life technologies MEGAscript T7 kit, using 1 µg of DraI digested plasmid as a template. After a 4 hour incubation, the sgRNA was purified by Ammonium Acetate precipitation per manufacturer's instructions.

Imaging

Imaging of *Pmyo-2::EGFP* and *Phsp-16.48::Cas9::EGFP* expressing animals was performed on an Andor Revolution spinning disc confocal microscope. Z-stacks with 1 µm slice distance were taken at several locations along the length of the worm. Stacks were then stitched together using

the ImageJ pairwise stitching plugin. Finally, a maximum intensity projection of 9 slices was generated.

Injections and heat shock induction

Plasmids and RNA were injected using standard *C. elegans* microinjection procedures. To induce expression from the *hps-16.48* promoter, injected animals were heat shocked for 1 hour at 34°C on agar plates floating in a water bath, 30 minutes to 1 hour after injection.

Reagent availability

The sgRNA and Cas9 expression plasmids will be made available through Addgene (http://www.addgene.org).

ACKNOWLEDGEMENTS

We thank M. Harterink for critical reading of the manuscript. Some strains were provided by the CGC, which is funded by the NIH Office of Research Infrastructure Programs (P40 OD010440). This work is supported by Innovational Research Incentives Scheme Vidi grant 864.09.008 and research grant 81902016, financed by the Netherlands Organization for Scientific Research (NWO/ALW).

REFERENCES

1. DiCarlo, J. E. *et al*. Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. Nucleic Acids Res. 1–8 (2013).
2. Gratz, S. J. *et al*. Genome engineering of *Drosophila* with the CRISPR RNA-guided Cas9 nuclease. Genetics 194, 1029–35 (2013).
3. Yu, Z. *et al*. Highly efficient genome modifications mediated by CRISPR/Cas9 in *Drosophila*. Genetics 195, 289–91 (2013).
4. Bassett, A. R., Tibbit, C., Ponting, C. P. & Liu, J.-L. Highly efficient targeted mutagenesis of *Drosophila* with the CRISPR/Cas9 system. Cell Rep. 4, 220–8 (2013).
5. Mali, P. *et al*. CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. Nat. Biotechnol. 31, 833–8 (2013).
6. Cho, S. W., Kim, S., Kim, J. M. & Kim, J.-S. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. Nat. Biotechnol. 3–5 (2013). doi:10.1038/nbt.2507
7. Hwang, W. Y. *et al*. Efficient genome editing in zebrafish using a CRISPR-Cas system. Nat. Biotechnol. 31, 227–9 (2013).
8. Chiu, H., Schwartz, H. T., Antoshechkin, I. & Sternberg, P. W. Transgene-free genome editing in *Caenorhabditis elegans* using CRISPR-Cas. Genetics 195, 1167–71 (2013).
9. Katic, I. & Großhans, H. Targeted heritable mutation and gene conversion by Cas9-CRISPR in *Caenorhabditis elegans*. Genetics 195, 1173–6 (2013).
10. Tzur, Y. B. *et al*. Heritable custom genomic modifications in *Caenorhabditis elegans* via a CRISPR-Cas9 system. Genetics 195, 1181–5 (2013).
11. Cho, S. W., Lee, J., Carroll, D., Kim, J.-S. & Lee, J. Heritable gene knockout in *Caenorhabditis elegans* by direct injection of Cas9-sgRNA ribonucleoproteins. Genetics 195, 1177–80 (2013).
12. Lo, T.-W. *et al*. Precise and heritable genome editing in evolutionarily diverse nematodes using TALENs and CRISPR/Cas9 to engineer insertions and deletions. Genetics 195, 331–48 (2013).
13. Jinek, M. *et al*. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337, 816–21 (2012).
14. Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proc. Natl. Acad. Sci. U. S. A. 109, E2579–86 (2012).

15. Bessereau, J. L. *et al*. Mobilization of a *Drosophila* transposon in the *Caenorhabditis elegans* germ line. Nature 413, 70–4 (2001).
16. Merritt, C., Rasoloson, D., Ko, D. & Seydoux, G. 3' UTRs are the primary regulators of gene expression in the *C. elegans* germline. Curr. Biol. 18, 1476–82 (2008).
17. Frøkjær-Jensen, C., Davis, M. W., Ailion, M. & Jorgensen, E. M. Improved Mos1-mediated transgenesis in *C. elegans*. Nat. Methods 9, 117–8 (2012).
18. Thomas, J., Lea, K., Zucker-Aprison, E. & Blumenthal, T. The spliceosomal snRNAs of *Caenorhabditis elegans*. Nucleic Acids Res. 18, 2633–42 (1990).
19. Friedland, A. E. *et al*. Heritable genome editing in *C. elegans* via a CRISPR-Cas9 system. Nat. Methods 10, 741–3 (2013).
20. Brenner, S. The genetics of *Caenorhabditis elegans*. Genetics 77, 71–94 (1974).
21. Redemann, S. *et al*. Codon adaptation-based control of protein expression in *C. elegans*. Nat. Methods 8, 250–2 (2011).
22. Imburgio, D., Rong, M., Ma, K. & McAllister, W. T. Studies of promoter recognition and start site selection by T7 RNA polymerase using a comprehensive collection of promoter variants. Biochemistry 39, 10419–30 (2000).
23. Fruscoloni, P., Zamboni, M., Panetta, G., De Paolis, A. & Tocchini-Valentini, G. P. Mutational analysis of the transcription start site of the yeast tRNA(Leu3) gene. Nucleic Acids Res. 23, 2914–8 (1995).
24. Zecherle, G. N., Whelen, S. & Hall, B. D. Purines are required at the 5' ends of newly initiated RNAs for optimal RNA polymerase III gene expression. Mol. Cell. Biol. 16, 5801–10 (1996).

SUPPORTING INFORMATION

Table S1. Concentration dependency of the embryonic lethality caused by *Peft-3::Cas9*.

	Transgenic F1		
***Peft-3::Cas9* conc.**	**Embryonic lethal**	**Viable**	**% Emb**
0 ng/µl	3	28	10
1 ng/µl	6	14	30
2 ng/µl	12	15	44
5 ng/µl	6	7	46
10 ng/µl	20	4	83
20 ng/µl	17	0	100

Injections consisted of 50 ng/µl of sgRNA, 5 ng/µl of the *Pmyo-3::mCherry* marker to identify transgenic animals, and the indicated amounts of *Peft-3::Cas9*. To inject a constant amount of DNA (75 ng/µl), injections with less than 20 ng/µl of *Peft-3::Cas9* were supplemented with empty pBluescript vector. Results represent the transgenic progeny derived from 6 injected animals over a 28 hour period.

Chapter 4

Engineering the *Caenorhabditis elegans* genome with CRISPR/Cas9

Selma Waaijers & Mike Boxem

Methods (2014) Mar 28; doi: 10.1016/j.ymeth.2014.03.024.

Utrecht University, Department of Biology, Padualaan 8, 3584 CH Utrecht, The Netherlands.

ABSTRACT

The development in early 2013 of CRISPR/Cas9-based genome engineering promises to dramatically advance our ability to alter the genomes of model systems at will. A single, easily produced targeting RNA guides the Cas9 endonuclease to a specific DNA sequence where it creates a double strand break. Imprecise repair of the break can yield mutations, while homologous recombination with a repair template can be used to effect specific changes to the genome. The tremendous potential of this system led several groups to independently adapt it for use in *Caenorhabditis elegans*, where it was successfully used to generate mutations and to create tailored genome changes through homologous recombination. Here, we review the different approaches taken to adapt CRISPR/Cas9 for *C. elegans*, and provide practical guidelines for CRISPR/Cas9-based genome engineering.

1. INTRODUCTION

The ability to engineer specific changes to the genome – deleting genes, replacing genes with mutant variants, or adding sequences encoding protein tags – is essential for researchers working with model organisms. For *Caenorhabditis elegans*, numerous methods for modifying the genome have been developed. These include the random introduction of mutations through chemical means, and the random insertion of transgenes in the genome by integration of extrachromosomal arrays, or as low copy insertions through microparticle bombardment.[1,2] More recently, methods based on mobilization of the *Mos1* transposon have enabled targeted genome engineering.[3,4] In this approach, mobilization of a pre-existing insertion of the *Drosophila Mos1* transposable element results in the generation of a double strand break (DSB). By offering a repair template containing sequences identical to the regions flanking the DSB, a sequence of choice can be inserted in the genome through homologous recombination. This approach, termed *Mos*TIC for '*Mos1* excision-induced transgene-instructed gene conversion,' was developed further to allow rapid single copy insertion of transgenes at several well-defined loci in the genome (*Mos*SCI) and to generate targeted deletions (*Mos*DEL).[5–7] However, only a limited fraction of the genome is accessible for *Mos1*-based genome engineering, as the efficiency of homologous-recombination declines rapidly with increased distance from the DSB site.[3] The ~13,300 *Mos1* insertions currently available to the community target ~40% of all *C. elegans* genes, and in many cases only a limited region of the gene can be targeted.[8]

The adaptation of the *Streptococcus pyogenes* type II CRISPR system to create targeted DSBs now promises to provide a simple and efficient tool with which the entire *C. elegans* genome is accessible for engineering. Studies into the mechanisms by which bacteria defend against bacteriophages led to the discovery of CRISPR adaptive immune systems, which target foreign DNA sequences for cleavage.[9–11] CRISPR stands for clustered regularly interspaced short palindromic repeats, which describes the specific repetitive makeup of the bacterial loci that initially attracted attention to them.[12] In the *S. pyogenes* CRISPR system, the endonuclease Cas9 is targeted to specific DNA sequences by two short non-coding RNA products: the CRISPR RNA (crRNA) which contains the 20 nt guide sequence that mediates target recognition, and the supporting trans-activating crRNA (tracrRNA). DNA cleavage by Cas9 additionally depends on the presence of a short consensus sequence (NGG) immediately adjacent to the target site, termed the protospacer adjacent motif, or PAM.[13,14] Absence of the PAM sequence in the CRISPR locus prevents cleavage of the host DNA. After target recognition, the coordinated action of two nuclease domains results in a double strand break 3 bp upstream of the PAM.[13,14]

A key step in developing CRISPR/Cas9 as a tool for genome engineering was the finding that *in vitro*, the Cas9 endonuclease can be targeted to specific sites by a single guide RNA (sgRNA), in which the 3' end of crRNA is fused to the 5' end of tracrRNA.[14] In January of 2013, five groups reported the adaptation of this system to mutate specific genes in mammalian cells and zebrafish.[15–19] Following these initial publications, adoption of this

system ballooned and by the end of 2013, CRISPR/Cas9 had been adapted for use in many model organisms including yeast, flies, zebrafish, mice, rat, plants, and *C. elegans*.[17,20–39]

The power of the CRISPR/Cas9 system lies in its simplicity and ease of use. To generate DSBs, only two components need to be supplied: the Cas9 protein, and the sgRNA. The targeting specificity is encoded in the first 20 nt of the sgRNA (Fig. 1), which makes it easy to engineer different sgRNAs targeting different sites. This is a distinct advantage compared to two other commonly used engineered nucleases, TALENs and zinc finger nucleases, for which a new protein coding sequence has to be generated for each DNA target. Finally, selection of Cas9 target sites is only limited by the presence of the GG dinucleotide PAM motif.

In this review we give an overview of the approaches taken to use CRISPR/Cas9 in *C. elegans*. Rather than presenting one specific protocol, we highlight different aspects of these approaches and provide practical guidelines to genome engineering in the worm. Individual protocols can be found in the original manuscripts covered in this article.[31–40]

2. ADAPTATION OF CRISPR/CAS9 FOR *C. ELEGANS*

Given the potential benefits, it is not surprising that several groups worked in parallel to adapt the CRISPR/Cas9 system for *C. elegans*. A team led by John Calarco was the first to demonstrate activity of Cas9 in *C. elegans*, using the system to generate mutations in several genes by non-homologous end joining.[35] This publication was rapidly followed by eight other reports of CRISPR/Cas9 usage in *C. elegans*, several of which also demonstrated homologous recombination-mediated gene conversion.[31–34,36–39] The main differences between the approaches lie in the methods employed to supply Cas9 protein and sgRNA to the germline of *C. elegans*.

2.1 Expression of Cas9

To obtain heritable changes in the genome, Cas9 protein and sgRNA need to be expressed in the germline of *C. elegans*. The methods used to supply Cas9 are summarized in Figure 2A. Most groups used a plasmid DNA-based approach, expressing *C. elegans* codon optimized Cas9 from the *eft-3* promoter and *tbb-2* 3'-UTR, a combination previously shown to be effective in expressing the *Mos1* transposase in the germline.[5,31,34,35,38,39] In addition, we expressed Cas9 from the heat-shock promoter *hsp-16.48*, which can be used to express Cas9 during a limited time window and

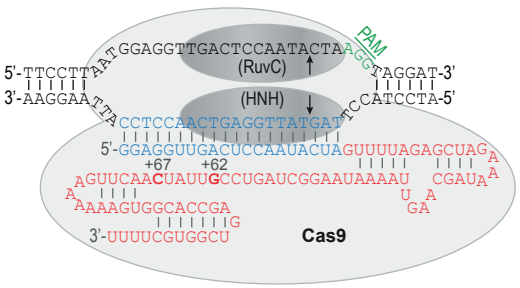


Figure 1. Example of Cas9/sgRNA in complex with a target site. The RuvC and HNH endonuclease domains of Cas9 together generate a double strand break. The PAM sequence in the genome is in green, the 20 nt recognition sequence in blue, and the remainder of the sgRNA sequence in red. Bold red letters with coordinates are numbered from the first base of the sgRNA, and indicate alternative end positions of the sgRNAs used by references 34 (+62) and 38 (+67).

circumvents the deleterious side-effects we observed after expressing Cas9 from the *eft-3* promoter.[39] As an alternative delivery method, three groups expressed Cas9 from injected *in vitro* synthesized capped and poly-adenylated mRNA, using both codon optimized and non-optimized versions with similar success rates.[32,36,37] Finally, direct injection of purified Cas9 protein was also shown to be an effective delivery method.[33] Injections of mRNA, and especially protein, may provide a higher dose of Cas9 to the germline, since no time is needed to allow transcription of the gene. In addition, similar to heat-shock promoter based expression, injection of mRNA should result in a short time window of Cas9 activity, dictated by mRNA and protein stability and the rate of translation. Such a window of activity was also observed for the generation of indels by injection of TALEN mRNA.[41] The *in vitro* production of mRNA and the purification or purchasing of Cas9 protein are more laborious and expensive than the use of plasmid DNA. Since both approaches were successful, the method of choice can be based on individual preferences.

2.2 Expression of sgRNA

To supply the sgRNA, again a choice exists between injecting *in vitro* transcribed sgRNA, or expressing the sgRNA from a plasmid (Fig. 2B). Three groups report the successful use of *in vitro* transcribed sgRNA.[32,33,36] Injection of RNA has the benefit that the sgRNA is directly supplied to the germline, and at very high concentrations if desired. Compared to the production of Cas9 mRNA, *in vitro* transcription of the short sgRNAs is a simpler and more efficient procedure, as the sgRNA sequence is short and does not need to be capped or poly-adenylated. Thus, high concentrations of sgRNA can readily be produced. Nevertheless, it appears that *in vivo* transcription of sgRNA is no less effective, as four publications report the successful use of a plasmid based approach where the sgRNA is placed under control of a promoter that normally drives expression of a non-coding RNA.[31,34,35,39] Each used a U6 RNA polymerase III promoter to drive expression of the sgRNA. Interestingly, the promoters of four different U6 genes were used, with the included genomic fragments ranging from 157 to 500 bp, all of which resulted in successful expression of sgRNA. In addition, the promoter of the *rpr-1* snRNA has been used successfully.[31]

One group used a somewhat different approach, injecting separate *in vitro* transcribed crRNA and tracrRNA, because a sgRNA was not functional *in vivo* and far less effective at target cleavage *in vitro*.[37] This likely resulted from the length of the sgRNA used in this study (+67 in Fig. 1), which lacked a 30 nucleotide 'tail' present in tracrRNA. Another study did report success with a shorter sgRNA (+62 in Fig. 1).[33] However, in human cells, addition of the 30 nt tail to the sgRNA increased the rate of generating indels, with up to 5-fold greater rates reported than the corresponding crRNA-tracrRNA duplex.[42] Thus, although a shorter sgRNA can be effective, the longer form appears most efficient and is preferred.

2.3 Choosing the best strategy

A direct comparison between the different approaches has not been made, and each group chose a different set of target genes and reports efficiencies in different ways. Moreover, certain combinations of reagents, *e.g.*, *in vitro* transcribed Cas9 combined with plasmid

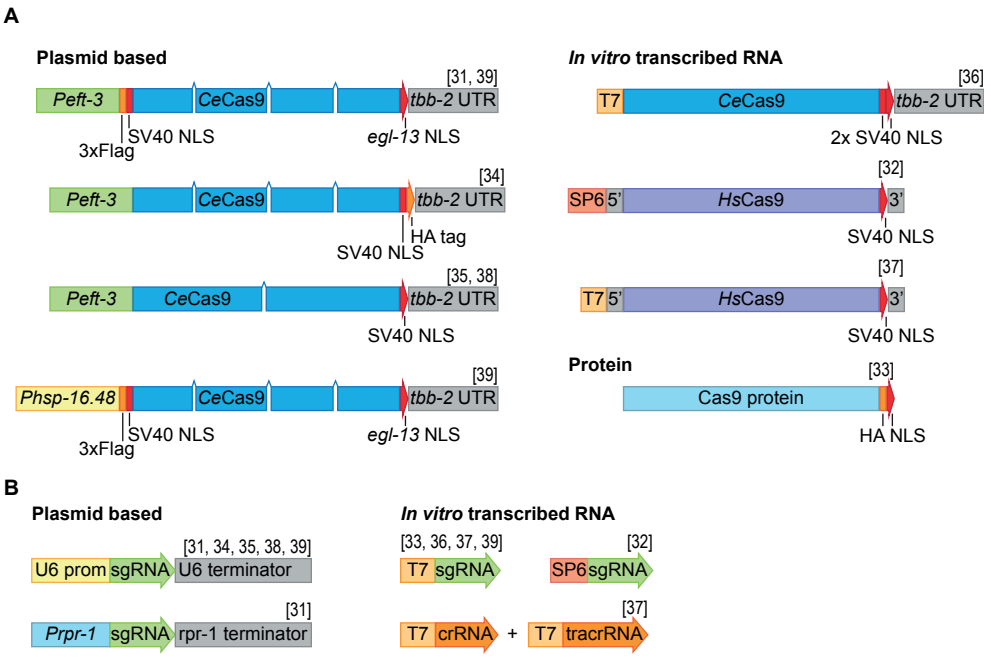


Figure 2. Schematic representation of the different Cas9 protein delivery methods (A) and sgRNA delivery methods (B) used by the publications discussed in this review. Drawings are not to scale. The different sizes of the 4 different U6 promoters used are not indicated separately. *Ce*Cas9 and *Hs*Cas9 indicate Cas9 constructs with *C. elegans* and human codon bias, respectively. References to the publications using each approach are indicated above the drawings.

encoded sgRNA, have not been tested. Nevertheless, the reported efficiencies are similar. For this reason, cost and ease of reagent generation make plasmid-based approaches the most attractive option for most purposes. A need for a more narrow window of CRISPR/Cas9 activity might favor the use of heat shock promoter driven Cas9, or RNA and protein injection-based approaches. The latter two may also exhibit a stronger activity, which may outweigh the added expense and complexity of reagent generation in particular cases.

3. TARGET SITE SELECTION

The first step in applying CRISPR/Cas9 is the selection of appropriate target sites. CRISPR/Cas9 targets different sites with variable efficiencies. The rules governing this variation, however, have not yet been discovered. For the moment, target sites are therefore selected based on only two criteria. First, the need for a PAM consensus sequence and the promoter used to express the sgRNA impose some restrictions on the sequence that can be targeted. Second, it is advisable to attempt to minimize the potential for off-target effects, although thus far in *C. elegans* this does not seem to pose a severe problem.

3.1 Sequence constraints

For efficient targeting, the three nucleotides following the target site need to adhere to the PAM sequence NGG. In addition, the approach used to provide the sgRNA may impose

constraints on the first two nucleotides of the sgRNA. For plasmid-based delivery, efficient transcription from RNA polymerase III reporters has been shown to benefit from the presence of a purine at the +1 position.[43,44] For *in vitro* transcription of mRNA, the commonly used T3 and T7 promoters heavily favor GG at the +1 and +2 positions, while the consensus sequence for SP6 at these positions is GA.[45–48] Incorporating these constraints results in consensus sequences of G/A-(N19)-NGG for U6 plasmid-based transcription, GG-(N18)-NGG for T3 and T7 *in vitro* transcription, and GA-(N18)-NGG for SP6 *in vitro* transcription. There are indications however that rules governing the 5' end of the sgRNA can be circumvented by adding additional nucleotides to the 5' end of the sgRNA. In *C. elegans*, a sgRNA that contained an extra GG pair at its 5' end was used successfully to generate mutations in *ben-1*.[36] In mammalian cells, extension of a sgRNA with 10 nucleotides had little effect on genome modification efficiency, even if the 8 nt immediately upstream of the 20 nt guide sequence did not match the target DNA.[49] The longer sgRNAs appeared to be processed to the same length as sgRNAs containing a 20 bp guide sequence, eliminating the extra bases.[49] These experiments indicate that any nucleotides required by the promoter used can be added upstream of the sgRNA, yielding N20-NGG as the consensus sequence.

A recent report tested the use of truncated sgRNAs to reduce off-target effects, and showed that sgRNAs with guide sequence lengths as short as 17 nt still resulted in efficient targeting.[50] As both shorter and longer sgRNA sequences appear to be functional, varying the sgRNA length can be used to design fully complementary sgRNAs that start with the desired nucleotide. A minimum size of 17 nt and a possible extension of at least 10 nt yields consensus sequences of G/A-N(16-29)-NGG, GG-N(15-28)-NGG, and GA-N(15-28)-NGG for U6, T3/T7, and SP6 promoters respectively. Reducing or eliminating constraints on the 5' sequence of the sgRNA would significantly increase the number of potential CRISPR/Cas9 target sites.

3.2 Off-target effects

A potential concern is the possibility of off-target effects, which could generate mutations at undesired sites. In human cells, CRISPR/Cas9 has been shown to cause cleavage at off-target sites.[42,51] The specificity of the Cas9/CRISPR complex depends on the position, number, and identity of the mismatched bases. Up to 5 mismatches have been reported to be tolerated, with mismatches close to the PAM less tolerated than mismatches more distal.[42,51] Two experimental strategies that reduce off-target effects have been reported in mammalian cells. The first is the use of a double nicking strategy based on a mutant Cas9 that generates a single strand nick. By targeting two Cas9 molecules to opposite strands of a target locus, DSBs are generated with minimal off-target effects.[49,52,53] The second strategy uses shortened sgRNAs, reducing the length of the guide sequence from 20 to 17 nt.[50] These truncated guide RNAs (tru-gRNAs) showed reduced mutagenesis at off-target sites without adversely affecting mutagenesis rates at on-target sites. The combination of these two approaches further reduced off-target effects.[50]

Several groups analyzed the mutant *C. elegans* strains they obtained, sequencing candidate off-target sites, and in one case even performing whole-genome sequencing.[32,34,35] None

observed mutations at sites other than the intended target. It appears therefore that the chance that CRISPR/Cas9 induces a mutation at an off-target site is too low for such an event to have occurred in the limited number of strains examined. Nevertheless, it remains advisable to follow the general rule that phenotypes should only be analyzed after backcrossing, and that multiple independent alleles should be analyzed if possible. In addition, computational methods can be used to avoid target sequences with a high chance of off-target effects (see section 3.3 below).

3.3 Computational approaches to target site selection

A simple approach to identifying suitable CRISPR target sites is to use a sequence editing software package that allows searching with wildcards. For example, in the freely available Ape – A plasmid editor, searching for RnnnnnnnnnnnnnnnnnnnGG will find any site matching the A/G-(N19)-NGG consensus.[54] Note that target sites can be present on either strand. More refined searches can be accomplished using computational tools that incorporate advanced features. For example, several web-based tools have been developed that assist in CRISPR target site selection and minimization of off-target effects.[42,55,56] Presumably, the number of such websites will increase and they will evolve to incorporate any future design rules that are discovered.

4. GENERATING MUTATIONS THROUGH NON-HOMOLOGOUS END-JOINING

DSB repair through non-homologous end-joining (NHEJ) is susceptible to the generation of small insertions or deletions (indels), which may disrupt gene function. The generation of such mutations requires only the targeting of a site with CRISPR/Cas9 followed by mutant selection (Fig. 3A). This approach is efficient in *C. elegans*, and numerous different loci were targeted for mutation in this manner.[31–33,35–37,39] In general, injection of 20 P0 animals should result in the identification of multiple mutant alleles.

One publication reported F1 animals that already harbored a homozygous mutation.[35] Most likely, a mutation acquired in the germline in one chromosome is copied to the second chromosome in the just fertilized embryo through homology-mediated repair. It is also possible, though unlikely, that both the oocyte and sperm derived chromosome independently acquire the same mutation by chance.

Most reported mutations fell within the expected spectrum of small indels. Two groups however reported a prevalence of larger deletions up to >3 kb in size.[32,33] In one case, purified Cas9 protein was used, which may cause a different mutation spectrum for unknown reasons.[33] Injection of *in vitro* transcribed Cas9 mRNA and sgRNA also resulted in larger deletions.[32] As a different group did not identify such deletions while using the same approach, it is not immediately clear how these differences arose.[36]

An interesting strategy to generate larger deletions may be the simultaneous targeting of two sites flanking the intended deletion (Fig. 3B). Several mutations were generated in the *dpy-3* gene using two sgRNAs that target two sites located 53 bp apart.[33] One of the deletions identified exactly spans the region predicted to be excised when Cas9 cuts at both sites. It

should be possible to use this approach to target entire genes for deletion, creating clean molecular null alleles. Such deletions could easily be identified in the F1 generation by PCR amplification.

4.1 Strategies for mutant selection

After injecting P0 animals with the Cas9 and CRISPR delivery method of choice, a number of convenient options exist to identify mutant progeny. First, if the phenotype caused by the mutation is known and easily identified, mutations can simply be identified by isolating F1 animals and examining their progeny for the expected phenotype. If the phenotype is not known or cannot easily be scored, several PCR-based approaches can be used to identify mutant F1 animals. The simplest is to sequence a PCR amplicon spanning the targeted site. A heterozygous indel will cause the sequence trace to degrade into two sets of overlapping peaks somewhere near the targeted site. This approach can also identify homozygous F1 animals, which have occasionally been observed to arise (see section 4 above).

A second approach is to analyze PCR amplicons spanning the targeted site for polymorphisms using an endonuclease that specifically recognizes and cleaves mismatched DNA, such as T7 endonuclease I, CEL I, Surveyor nuclease, or mung bean nuclease.[57–60] In this approach, the PCR amplicon is denatured and allowed to re-anneal. Heterozygous mutations will result in a mismatch in approximately half of the re-annealed product, which will be recognized and cleaved by the endonuclease. Since homozygous mutations do not result in a mismatch, these would not be recognized by this approach. Detecting mutations using endonucleases has two advantages over sequencing. First, the approach can be more cost effective, especially when using homemade CEL I or mung bean nuclease, which can be isolated from celery and mung bean sprouts respectively.[60,61] Second, the time to identification of mismatches is reduced, since the endonuclease reactions can be performed in-house the same day the PCR products are generated.

As a final approach, the sgRNA target sequence can be chosen to overlap with a restriction site, although this severely limits the target site choice. Hetero- or homozygous mutations resulting in loss of the restriction site can be detected as a restriction fragment length polymorphism after PCR amplification and digestion. Appropriate controls have to be used to ensure that an observed lack of digestion is not due to technical issues.

Most reports focused their analysis on transgenic F1 animals, identified by expression of a co-injected fluorescent marker protein. Of two studies that also examined non-transgenic F1 progeny, one recovered mutations from both transgenic and non-transgenic animals, while the second identified no mutations among the non-transgenic F1 animals.[31,35] Thus, transgenic animals may have a higher chance of carrying mutations.

A drawback of the PCR-based approaches is that larger deletions spanning one or both of the primer binding sites will not be uncovered. Furthermore, it is important to take into account that CRISPR/Cas9 likely can result in somatic mutation events. In one published report, F1 animals were observed with more than two mutations in the *unc-1* gene, as well as an F1 animal with the Unc-1 phenotype that nevertheless contained both wild-type

and mutant *unc-1* sequences.[33] These observations indicate the occurrence of mutations in somatic cells, which are present in the F1 generation but are not transmitted to the F2 progeny. In most cases however, multiple candidate mutants will be identified in the F1, which can be verified as true heritable germline events by examining the F2.

5. HOMOLOGOUS RECOMBINATION-BASED GENOME ENGINEERING

One of the most exciting applications of CRISPR/Cas9 is the direct engineering of the genome, modifying genes and adding specific tags through homology mediated repair of the double strand break. By supplying a repair template containing sequences identical to the DNA flanking the break site, any sequence of choice can be inserted into the genome (Figure 3E-G). If the homologous regions are further apart in the genome, endogenous sequences can also be replaced with a desired sequence, for example to create a deletion mutant (Figure 3C and D) or to replace a gene with a specific mutant variant (Figure 3H-J). Most commonly, genome editing through homologous recombination makes use of plasmid-based repair templates carrying homologous arms of varying lengths. A number of usage scenarios have been demonstrated in *C. elegans* using this approach. First, GFP expressing transgenes were inserted near a *MosI* insertion site and into the *klp-12* locus, and the HygR gene – which confers resistance to hygromycin B – was inserted at two different sites in the *ben-1* locus.[31,34,38] Targeting a site near a *MosI* locus allowed a direct comparison of the efficiencies of CRISPR/Cas9 and the *Mos*SCI technique, which showed that the overall efficiency of inserting the same GFP expression cassette was comparable between the two approaches.[34] Second, the ability to delete an entire gene was illustrated by the replacement of the *lab-1* coding sequences with a GFP expression construct.[38] Third, the engineering of specific gene mutations was illustrated by the repair of a missense mutation in the *daf-2* gene, and the replacing of the *lin-31* gene with a mutant version lacking four MAPK phosphorylation sites in the corresponding protein.[34,36] Finally, the applicability of the system to tag endogenous genes was elegantly demonstrated by the addition of a GFP tag to the *nmy-2* and *his-72* genes.[34] The addition of the tag had no discernable negative effects on gene function, and the GFP tagged protein products faithfully recapitulated their known localization pattern. Although again the efficiency of inducing homologous recombination is difficult to compare between groups, overall, recombinant lines were identified from 2 – 20% of injected P0 animals.

Plasmid-based repair templates can be used to engineer almost any genome change imaginable. However, cloning of the repair template can be time consuming. For small modifications, such as changes of one or a few amino acids, a single strand oligodeoxynucleotide (ssODN) repair template can also be used, and may in fact be more efficient than plasmid-based repair templates.[62] ssODNs can also be used to generate deletions, by designing a template where the left and right halves of the ssODN flank the desired deletion.[62] The ease of obtaining oligonucleotides makes this an attractive alternative for small changes, with the caveat that this approach cannot be combined with positive selection (see below). ssODNs have recently been shown to be effective templates to engineer small changes in *C. elegans*, for both CRISPR/Cas9 and TALEN induced DSBs.[37,40] Interestingly, for ssODN mediated

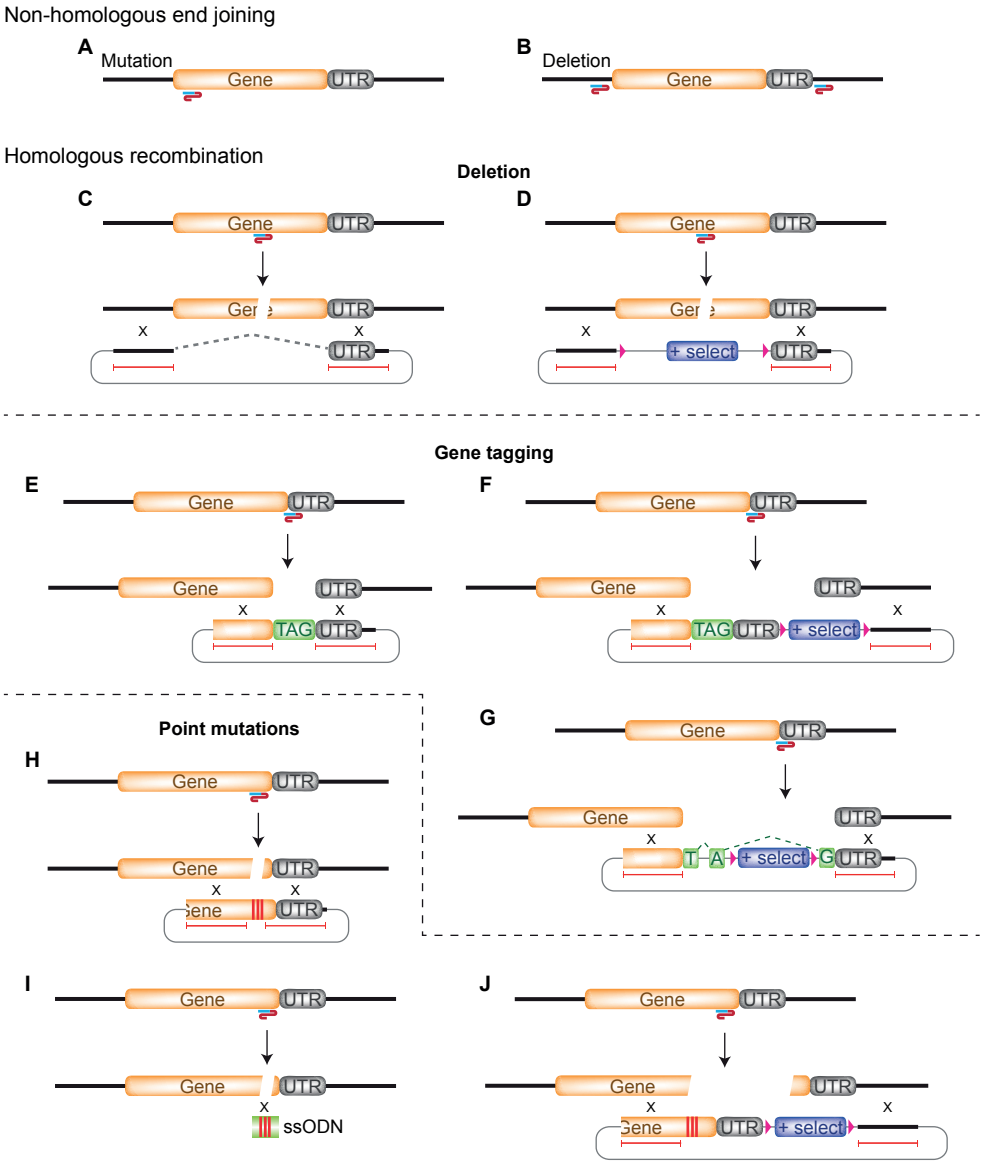


Figure 3. Examples of different genome engineering applications of CRISPR/Cas9. Target sites are indicated by the stylistic sgRNA. (A-B) Non-homologous end joining can be used to generate mutations and deletions. (C-J) Use of homologous recombination to generate deletions, insert gene tags, or engineer point mutations. Homologous regions are indicated by the red lines in the repair constructs. Pink triangles represent LoxP sites. For gene tagging, re-cleavage after repair is prevented by having the inserted tag split the target sequence. For the introduction of point mutations, re-cleavage is prevented by simultaneously mutating the PAM sequence.

repair of CRISPR/Cas9 induced DSBs, the desired changes were found almost exclusively in non-transgenic F1 animals.[40] The authors hypothesize that this is likely due to continued expression of active Cas9 in transgenic animals, which would lead to re-cleavage of already correctly engineered DNA.

5.1 Design of the repair template

A repair template consists of the desired genome modification flanked by homologous arms that are identical in sequence to the target DNA. For plasmid-based repair templates, the optimal length of the homologous arms has not yet been accurately determined. Early studies on homologous recombination in mouse embryonic stem cells showed an exponential increase in recombination frequency with increased length of the homologous region, up to ~14 kb of total homology.[63,64] In contrast, studies on *MosI*-mediated gene conversion in *C. elegans* indicated that arm lengths greater than 1.5 kb each do not improve repair efficiency. Shorter arm lengths are rarely used, although a length of 1.0 kb for each arm has been shown to successfully mediate CRISRP/Cas9-induced DSBs.[38] As arm lengths of 1.5 kb efficiently repair DSBs introduced by *MosI* excision and CRISRP/Cas9, this represents a safe length to use in repair template design. For single strand oligodeoxynucleotide (ssODN) repair templates, the optimal length of the homologous arms in mammalian cells is ~45 nt for each arm, with both shorter and longer arms reducing repair frequency.[62,65] The reduced efficiency in mammalian cells seen with longer arms may be due to cell culture specific toxicity effects, and it would be worthwhile to repeat this analysis in *C. elegans*. Studies using ssODNs in *C. elegans* successfully used homologous arm lengths of 20 and ~50 nt.[37,40] ssODNs can be designed to match either strand of the target DNA, although ssODNs that are complementary to the sgRNA resulted in higher repair efficiencies in mammalian cells.[65]

For both plasmid-based and ssODN repair templates, the efficiency of recombination decreases rapidly with distance from the DSB site.[3,65] The most efficient template design should therefore place the homologous arms on either side of the Cas9 target sequence. Furthermore, if a positive selection marker is included in the template (see section 5.2 below), it is important to design the repair template such that the selectable marker and the Cas9 target site flank the genomic region to be engineered. This ensures that any homologous recombination events that incorporate the selectable marker also include the desired genome modifications.

It is important that the repair template does not include the Cas9 recognition site. Presence of the recognition site in the template poses two problems. First, the repair template itself may be cleaved, rendering it unsuitable for homologous recombination-based repair. Second, after homologous recombination has taken place, the newly engineered genome can be cleaved, leading to unintended mutations by non-homologous end-joining. To eliminate the Cas9 recognition site from the repair template, one option is to mutate the PAM sequence, which is essential for cleavage. If this is not feasible, the 20 nt recognition sequence itself can be mutated. Due to the possibility of off-target cleavage, a minimum of 5 nucleotides should be changed, preferably close to the 3' end of the sequence as this region appears most important for target site selection.[42,51] For recombination events where

an exogenous sequence like a fluorescent protein tag is inserted, it can also be convenient to have the Cas9 target sequence span the exact insertion site. In the repair template, the target site will be split by the sequence to be inserted. However, this approach will not always be possible due to target site sequence constraints (see section 3.1). For ssODN repair templates, sequence restrictions may make it impossible to eliminate the Cas9 recognition site from the repair template. In this case, the chances of re-cleavage can be minimized by using a Cas9 delivery method that only results in a short window of activity (heat shock promoter or mRNA injections), or by focusing on non-transgenic F1 animals that did not incorporate the Cas9 expression plasmid into an extrachromosomal array.[40]

5.2 Identification of recombinants

Successful recombinants can be identified through a PCR based approach. If the recombination event introduces a unique primer binding site, the amplification reaction itself can be used to identify successful recombinants. One primer is designed to anneal to the newly introduced priming site, while the second primer is chosen outside of the sequences present in the repair construct. In this fashion, neither the endogenous locus nor the repair plasmid itself can be used as a template. In many cases, the insertion of exogenous sequences automatically creates an appropriate primer binding site, but even if only a few basepairs are to be changed, additional silent base pair changes can be incorporated to create a unique priming site. An alternative approach is to generate or remove a restriction enzyme site, and identify recombinants by digestion of a PCR product with appropriate enzymes. Finally, the target region can be PCR amplified and sequenced.

If homologous recombination is efficient enough, these approaches can be used to directly identify recombinants without any form of preselection, by analyzing individual F1 progeny of injected animals. Again, the risk of somatic events exists, and recombination events should be confirmed in subsequent generations. Alternatively, recombinants can first be selected through the use of positive and negative selection markers, before confirmation by PCR amplification and sequencing. A major benefit of direct identification is that no additional sequences need to be such as removal of a selection marker, is needed. This approach may also be the only option for ssODN mediated repair, where inclusion of a positive selection marker is not possible.

To identify recombinants based on selection markers, transgenic F1 animals are isolated and allowed to produce F2 or F3 progeny. Successful recombination events are then identified by a combination of positive selection for a marker present in the repair construct, and negative selection to eliminate animals that carry an extrachromosomal array. Correct recombination is confirmed by PCR amplification and sequencing of the target region.

The most commonly used positive selection markers are rescue of an *unc-119* mutation, and providing resistance against an antibiotic. Rescue of *unc-119* is commonly used in the *Mos*SCI method, where it has proven to be a reliable selection marker.[5,6] Similarly, *unc-119* rescue was shown to work well for Cas9 induced homologous recombination.[34] A drawback is that experiments need to be carried out in an *unc-119* mutant strain, reducing the flexibility of the approach. For antibiotic resistance, three cassettes are available, providing resistance to

neomycin, puromycin, and hygromycin B.[66–68] Of these three antibiotics, *C. elegans* appears to be most sensitive to hygromycin B, allowing for a strong selection for animals carrying the *HygR* resistance cassette.[68] This cassette was successfully used to select animals with a transgene inserted by Cas9 induced homologous recombination.[31]

A potential drawback of the use of a positive selection marker is that insertion of the marker in the genome may interfere with expression of the targeted gene. Care should be taken therefore to place the marker sequences in a location that does not affect gene expression, for example downstream of the 3' UTR. An elegant approach to reduce the risk of interfering with normal gene regulation is to remove the marker after the selection procedure is completed. Flanking the marker region with *loxP* sites allows it to be removed by subsequent expression of the Cre recombinase, which leaves behind only a single *loxP* site after excision. Removal of the *unc-119*(+) cassette in this manner was shown to be highly efficient.[34] When the genome modification involves insertion of a protein tag containing introns (e.g., GFP), a convenient variation on this theme is to place the *loxP* flanked positive selection marker within an intron (Fig. 3G). This simplifies cloning of the repair construct, while the single *loxP* site that is left after Cre induced recombination is unlikely to affect gene expression.

When a positive selection marker is used, it will be necessary to distinguish between expression of the marker from a genome-integrated copy and expression from an extrachromosomal array. This generally entails eliminating animals that carry an extrachromosomal array following the positive selection step. The most common approach is to look for absence of the co-injection marker used. In addition, it is possible to co-inject a counter selectable marker, such as heat-shock driven expression of the PEEL-1 toxin, which efficiently kills *C. elegans* upon ectopic expression.[5,69]

Although the above approaches should be generally applicable, other strategies may be more efficient for specific usage cases. For example, homologous recombination can be detected visually if the induced change causes a visible phenotype, such as expression of a fluorescent protein, or induction/reversal of a mutant phenotype.[31,36,38] This approach is especially powerful if an extrachromosomal array carrying the repair template cannot cause the phenotype.

6. CONCLUDING REMARKS

The publications reviewed here represent the first applications of CRISPR/Cas9 in *C. elegans*. Different groups used different approaches to providing Cas9 and sgRNA, and a direct comparison between approaches has not been made. A question that remains therefore is whether there is a single 'best' approach, or whether different approaches will prove to have specific advantages for certain uses. For example, homologous recombination may benefit from prolonged expression of Cas9, while generating mutations may be efficient enough that prolonged expression only increases the chances of off-target effects. In addition, the rules that govern the efficiency of targeting particular sequences are currently unknown. It is likely that future studies will address some of these questions and result in the development of improved protocols.

Applications of CRISPR/Cas9 in *C. elegans* have thus far been limited to the generation of mutants and homologous recombination-based genome editing. Impressive as these advances already are, the potential of the CRISPR/Cas9 system extends far beyond genome engineering. In other systems, a mutant version of Cas9 lacking endonuclease activity has already been used to modify transcriptional activity by targeting activators and repressors to specific DNA sequences, and to visualize specific genomic loci by targeting of EGFP.[53,70–77] In another recent development, RNA polymerase II driven production of sgRNA offers the possibility of applying CRISPR/Cas9 in a tissue-specific manner.[78] Conditional targeting of TALENs in somatic tissues has already been demonstrated in *C. elegans*, making it likely that it will be possible to use CRISPR/Cas9 to examine the effects of mutations in specific cell types.[79] Finally, CRISPR/Cas9 has made it possible for the first time to perform genetic mutagenesis screens in diploid human tissue culture cells.[80,81] The application of CRISPR/Cas9 is one of the most rapidly moving fields of research, and undoubtedly many exciting new applications will find their way to *C. elegans*.

ACKNOWLEDGEMENTS

We thank S. van den Heuvel and A. Thomas for critical reading of the manuscript. M.B. is supported by Innovational Research Incentives Scheme Vidi grant 864.09.008, financed by the Netherlands Organization for Scientific Research (NWO/ALW).

REFERENCES

1. Brenner, S. The genetics of *Caenorhabditis elegans*. Genetics 77, 71–94 (1974).
2. Praitis, V. & Maduro, M. F. Transgenesis in *C. elegans*. Methods Cell Biol. 106, 161–85 (2011).
3. Robert, V. & Bessereau, J.-L. Targeted engineering of the *Caenorhabditis elegans* genome following *Mos1*-triggered chromosomal breaks. EMBO J. 26, 170–83 (2007).
4. Robert, V. J. P., Katic, I. & Bessereau, J.-L. *Mos1* transposition as a tool to engineer the *Caenorhabditis elegans* genome by homologous recombination. Methods 49, 263–9 (2009).
5. Frøkjær-Jensen, C., Davis, M. W., Ailion, M. & Jorgensen, E. M. Improved *Mos1*-mediated transgenesis in *C. elegans*. Nat. Methods 9, 117–8 (2012).
6. Frøkjaer-Jensen, C. *et al*. Single-copy insertion of transgenes in *Caenorhabditis elegans*. Nat. Genet. 40, 1375–83 (2008).
7. Frøkjaer-Jensen, C. *et al*. Targeted gene deletions in *C. elegans* using transposon excision. Nat. Methods 7, 451–3 (2010).
8. Vallin, E. *et al*. A genome-wide collection of *Mos1* transposon insertion mutants for the *C. elegans* research community. PLoS One 7, (2012).
9. Horvath, P. & Barrangou, R. CRISPR/Cas, the immune system of bacteria and archaea. Science 327, 167–70 (2010).
10. Bhaya, D., Davison, M. & Barrangou, R. CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. Annu. Rev. Genet. 45, 273–97 (2011).
11. Sorek, R., Lawrence, C. M. & Wiedenheft, B. CRISPR-mediated adaptive immune systems in bacteria and archaea. Annu. Rev. Biochem. 82, 237–66 (2013).
12. Jansen, R., Embden, J. D. A. van, Gaastra, W. & Schouls, L. M. Identification of genes that are associated with DNA repeats in prokaryotes. Mol. Microbiol. 43, 1565–75 (2002).
13. Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proc. Natl. Acad. Sci. U. S. A. 109, E2579–86 (2012).
14. Jinek, M. *et al*. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337, 816–21 (2012).

15. Cho, S. W., Kim, S., Kim, J. M. & Kim, J.-S. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. Nat. Biotechnol. 3–5 (2013). doi:10.1038/nbt.2507
16. Cong, L. *et al*. Multiplex genome engineering using CRISPR/Cas systems. Science 339, 819–23 (2013).
17. Hwang, W. Y. *et al*. Efficient genome editing in zebrafish using a CRISPR-Cas system. Nat. Biotechnol. 31, 227–9 (2013).
18. Jinek, M. *et al*. RNA-programmed genome editing in human cells. Elife 2, (2013).
19. Mali, P. *et al*. RNA-guided human genome engineering via Cas9. Science 339, 823–6 (2013).
20. DiCarlo, J. E. *et al*. Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. Nucleic Acids Res. 1–8 (2013). doi:10.1093/nar/gkt135
21. Gratz, S. J. *et al*. Genome engineering of *Drosophila* with the CRISPR RNA-guided Cas9 nuclease. Genetics 194, 1029–35 (2013).
22. Yu, Z. *et al*. Highly efficient genome modifications mediated by CRISPR/Cas9 in *Drosophila*. Genetics 195, 289–91 (2013).
23. Bassett, A. R., Tibbit, C., Ponting, C. P. & Liu, J.-L. Highly efficient targeted mutagenesis of *Drosophila* with the CRISPR/Cas9 system. Cell Rep. 4, 220–8 (2013).
24. Hwang, W. Y. *et al*. Heritable and precise zebrafish genome editing using a CRISPR-Cas system. PLoS One 8, (2013).
25. Li, D. *et al*. Heritable gene targeting in the mouse and rat using a CRISPR-Cas system. Nat. Biotechnol. 31, 681–683 (2013).
26. Shen, B. *et al*. Generation of gene-modified mice via Cas9/RNA-mediated gene targeting. Cell Res. 23, 720–3 (2013).
27. Zhou, J. *et al*. One-step generation of different immunodeficient mice with multiple gene modifications by CRISPR/Cas9 mediated genome engineering. Int. J. Biochem. Cell Biol. 46C, 49–55 (2013).
28. Li, W., Teng, F., Li, T. & Zhou, Q. Simultaneous generation and germline transmission of multiple gene mutations in rat using CRISPR-Cas systems. Nat. Biotechnol. 31, 684–6 (2013).
29. Ma, Y. *et al*. Generating rats with conditional alleles using CRISPR/Cas9. Cell Res. 24, 122–5 (2014).
30. Belhaj, K., Chaparro-Garcia, A., Kamoun, S. & Nekrasov, V. Plant genome editing made easy: targeted mutagenesis in model and crop plants using the CRISPR/Cas system. Plant Methods 9, (2013).
31. Chen, C., Fenk, L. A. & de Bono, M. Efficient genome editing in *Caenorhabditis elegans* by CRISPR-targeted homologous recombination. Nucleic Acids Res. 41, e193 (2013).
32. Chiu, H., Schwartz, H. T., Antoshechkin, I. & Sternberg, P. W. Transgene-free genome editing in *Caenorhabditis elegans* using CRISPR-Cas. Genetics 195, 1167–71 (2013).
33. Cho, S. W., Lee, J., Carroll, D., Kim, J.-S. & Lee, J. Heritable gene knockout in *Caenorhabditis elegans* by direct injection of Cas9-sgRNA ribonucleoproteins. Genetics 195, 1177–80 (2013).
34. Dickinson, D. J., Ward, J. D., Reiner, D. J. & Goldstein, B. Engineering the *Caenorhabditis elegans* genome using Cas9-triggered homologous recombination. Nat. Methods 10, 1028–34 (2013).
35. Friedland, A. E. *et al*. Heritable genome editing in *C. elegans* via a CRISPR-Cas9 system. Nat. Methods 10, 741–3 (2013).
36. Katic, I. & Großhans, H. Targeted heritable mutation and gene conversion by Cas9-CRISPR in *Caenorhabditis elegans*. Genetics 195, 1173–6 (2013).
37. Lo, T.-W. *et al*. Precise and heritable genome editing in evolutionarily diverse nematodes using TALENs and CRISPR/Cas9 to engineer insertions and deletions. Genetics 195, 331–48 (2013).
38. Tzur, Y. B. *et al*. Heritable custom genomic modifications in *Caenorhabditis elegans* via a CRISPR-Cas9 system. Genetics 195, 1181–5 (2013).
39. Waaijers, S. *et al*. CRISPR/Cas9-targeted mutagenesis in *Caenorhabditis elegans*. Genetics 195, 1187–91 (2013).
40. Zhao, P., Zhang, Z., Ke, H., Yue, Y. & Xue, D. Oligonucleotide-based targeted gene editing in *C. elegans* via the CRISPR/Cas9 system. Cell Res. 24, 247–50 (2014).
41. Wood, A. J. *et al*. Targeted genome editing across species using ZFNs and TALENs. Science 333, 307 (2011).

42. Hsu, P. D. *et al*. DNA targeting specificity of RNA-guided Cas9 nucleases. Nat. Biotechnol. 31, 827–32 (2013).
43. Fruscoloni, P., Zamboni, M., Panetta, G., De Paolis, A. & Tocchini-Valentini, G. P. Mutational analysis of the transcription start site of the yeast tRNA(Leu3) gene. Nucleic Acids Res. 23, 2914–8 (1995).
44. Zecherle, G. N., Whelen, S. & Hall, B. D. Purines are required at the 5' ends of newly initiated RNAs for optimal RNA polymerase III gene expression. Mol. Cell. Biol. 16, 5801–10 (1996).
45. Breaker, R. R., Banerji, A. & Joyce, G. F. Continuous *in vitro* evolution of bacteriophage RNA polymerase promoters. Biochemistry 33, 11980–6 (1994).
46. Milligan, J. F. & Uhlenbeck, O. C. Synthesis of small RNAs using T7 RNA polymerase. Methods Enzymol. 180, 51–62 (1989).
47. Brown, J. E., Klement, J. F. & McAllister, W. T. Sequences of three promoters for the bacteriophage SP6 RNA polymerase. Nucleic Acids Res. 14, 3521–6 (1986).
48. Stump, W. T. & Hall, K. B. SP6 RNA polymerase efficiently synthesizes RNA from short double-stranded DNA templates. Nucleic Acids Res. 21, 5480–4 (1993).
49. Ran, F. A. *et al*. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. Cell 154, 1380–9 (2013).
50. Fu, Y., Sander, J. D., Reyon, D., Cascio, V. M. & Joung, J. K. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. Nat. Biotechnol. (2014). doi:10.1038/nbt.2808
51. Fu, Y. *et al*. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nat. Biotechnol. 31, 822–6 (2013).
52. Cho, S. W. *et al*. Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. Genome Res. 24, 132–41 (2014).
53. Mali, P. *et al*. CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. Nat. Biotechnol. 31, 833–8 (2013).
54. Davis, M. W. ApE - A plasmid editor. at <http://biologylabs.utah.edu/jorgensen/wayned/ape/>
55. Heigwer, F., Kerr, G. & Boutros, M. E-CRISP: fast CRISPR target site identification. Nat. Methods 11, 122–3 (2014).
56. Bae, S., Park, J. & Kim, J.-S. Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. Bioinformatics (2014). doi:10.1093/bioinformatics/btu048
57. Babon, J. J., McKenzie, M. & Cotton, R. G. The use of resolvases T4 endonuclease VII and T7 endonuclease I in mutation detection. Methods Mol. Biol. 152, 187–99 (2000).
58. Oleykowski, C. A., Bronson Mullins, C. R., Godwin, A. K. & Yeung, A. T. Mutation detection using a novel plant endonuclease. Nucleic Acids Res. 26, 4597–4602 (1998).
59. Qiu, P. *et al*. Mutation detection using Surveyor nuclease. Biotechniques 36, 702–7 (2004).
60. Till, B. J., Burtner, C., Comai, L. & Henikoff, S. Mismatch cleavage by single-strand specific nucleases. Nucleic Acids Res. 32, 2632–41 (2004).
61. Yang, B. *et al*. Purification, cloning, and characterization of the CEL I nuclease. Biochemistry 39, 3533–41 (2000).
62. Chen, F. *et al*. High-frequency genome editing using ssDNA oligonucleotides with zinc-finger nucleases. Nat. Methods 8, 753–5 (2011).
63. Hasty, P., Rivera-Pérez, J. & Bradley, A. The length of homology required for gene targeting in embryonic stem cells. Mol. Cell. Biol. 11, 5586–91 (1991).
64. Deng, C. & Capecchi, M. R. Reexamination of gene targeting frequency as a function of the extent of homology between the targeting vector and the target locus. Mol. Cell. Biol. 12, 3365–71 (1992).
65. Yang, L. *et al*. Optimization of scarless human stem cell genome editing. Nucleic Acids Res. 41, 9049–61 (2013).
66. Giordano-Santini, R. *et al*. An antibiotic selection marker for nematode transgenesis. Nat. Methods 7, 1–6 (2010).
67. Semple, J. I., Garcia-Verdugo, R. & Lehner, B. Rapid selection of transgenic *C. elegans* using antibiotic resistance. Nat. Methods 7, 725–7 (2010).

68. Radman, I., Greiss, S. & Chin, J. W. Efficient and rapid *C. elegans* transgenesis by bombardment and hygromycin B selection. PLoS One 8, (2013).
69. Seidel, H. S. *et al*. A novel sperm-delivered toxin causes late-stage embryo lethality and transmission ratio distortion in *C. elegans*. PLoS Biol. 9, (2011).
70. Chen, B. *et al*. Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. Cell 155, 1479–91 (2013).
71. Gilbert, L. A. *et al*. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. Cell 154, 442–51 (2013).
72. Cheng, A. W. *et al*. Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system. Cell Res. 23, 1163–71 (2013).
73. Hu, J. *et al*. Direct activation of human and mouse Oct4 genes using engineered TALE and Cas9 transcription factors. Nucleic Acids Res. (2014). doi:10.1093/nar/gku109
74. Qi, L. S. *et al*. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell 152, 1173–83 (2013).
75. Maeder, M. L. *et al*. CRISPR RNA-guided activation of endogenous human genes. Nat. Methods 10, 977–9 (2013).
76. Perez-Pinera, P. *et al*. RNA-guided gene activation by CRISPR-Cas9-based transcription factors. Nat. Methods 10, 973–6 (2013).
77. Bikard, D. *et al*. Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. Nucleic Acids Res. 41, 7429–37 (2013).
78. Gao, Y. & Zhao, Y. Self-processing of ribozyme-flanked RNAs into guide RNAs *in vitro* and *in vivo* for CRISPR-mediated genome editing. J. Integr. Plant Biol. (2013). doi:10.1111/jipb.12152
79. Cheng, Z. *et al*. Conditional targeted genome editing using somatically expressed TALENs in *C. elegans*. Nat. Biotechnol. 31, 934–7 (2013).
80. Shalem, O. *et al*. Genome-scale CRISPR-Cas9 knockout screening in human cells. Science 343, 84–7 (2014).
81. Wang, T., Wei, J. J., Sabatini, D. M. & Lander, E. S. Genetic screens in human cells using the CRISPR-Cas9 system. Science 343, 80–4 (2014).

Chapter 5

The conserved *Caenorhabditis elegans* Crumbs protein family plays a nonessential role in epithelial polarity

Selma Waaijers, Elisabeth Kruse & Mike Boxem

Utrecht University, Department of Biology, Padualaan 8, 3584 CH Utrecht, The Netherlands.

ABSTRACT

Crumbs proteins are essential regulators of epithelial polarity in many animal systems. For the two Crumbs homologs of *Caenorhabditis elegans*, however, no essential role has been uncovered. Here, we identified and characterized an additional Crumbs family member in *C. elegans*, which we termed CRB-3 based on its similarity in size and sequence to mammalian CRB3. A translational CRB-3::GFP localized apical in the epithelial cells of the pharynx and the intestine. To identify the function of the three Crumbs family members in *C. elegans* development, we generated a triple Crumbs deletion mutant by sequentially removing the entire coding sequence for each gene using a novel CRISPR/Cas9 approach. Remarkably, animals lacking all three Crumbs homologs were viable. Overexpression of all three *C. elegans* Crumbs homologs induces an altered localization of the polarity protein PAR-3. These results indicate that the *C. elegans* Crumbs family members play a non-essential role in polarity establishment.

INTRODUCTION

Cell polarity is of vital importance for the proper development and functioning of epithelial tissues. Epithelial cells are polarized into distinct apical and basolateral plasma membrane domains. These two domains are separated by the apical junctional complex (AJC). The AJC provides a strong mechanical connection between cells. In addition, it creates a seal between cells that prevents passage of molecules through the paracellular space and diffusion of proteins and lipids between apical and basolateral domains. Next to the junctional proteins, three evolutionarily conserved groups of protein have been identified that control the establishment and maintenance of apical and basolateral membrane domains. The Scribble group (SCRIB/DLG/LGL) localizes to the basolateral side and establishes basolateral identity, while the apically localized PAR proteins (PAR-3/PAR-6/aPKC) and Crumbs complex (CRB/PALS1/LIN-7/PATJ) define apical identity.[1]

Crumbs, the central component of the Crumbs complex, was identified in a large scale *Drosophila melanogaster* screen for mutations affecting the morphology of the cuticle.[2] Proper functioning of *crumbs* is essential for the establishment of an apical membrane domain in epithelia and for the assembly of the AJC, and thereby for epithelial integrity.[3–7] Crumbs is a transmembrane protein that localizes to the apical membrane of epithelia, and concentrates just apical of the AJC.[3,4,8] Its large extracellular domain is composed of 29 epidermal growth factor (EGF)-like repeats and three Laminin G-like domains; its short intracellular region of 37 amino acids contains a FERM (4.1 protein/Ezrin/Radixin/Moesin)-binding domain and a C-terminal PDZ (PSD-95/Discs large/ZO-1)-binding motif (Fig. 1A).[9] The PDZ-binding motif mediates binding of Crumbs to the membrane-associated guanylate kinases (MAGUK) protein Stardust (Sdt, known as PALS1 in mammals).[10,11] In turn, Sdt binds to the other two partners of the core Crumbs complex, Lin-7 and PATJ, both of which contain one or more PDZ domains, and a Lin-2/Lin-7 (L27) domain which mediates binding to Sdt.[12–15] Interestingly, the small intracellular domain appears to mediate much of the functioning of Crumbs, as expression of only the intracellular domain coupled to a transmembrane domain is sufficient to rescue most of the phenotypes observed in a *crumbs* mutant fly.[5,9]

The members of the Crumbs complex are highly conserved in vertebrates. Crumbs itself is represented in vertebrates by three distinct family members: CRB1, CRB2, and CRB3. CRB1 and CRB2 structurally resemble *Drosophila* Crumbs, in that they consist of a large extracellular domain made up of EGF repeats and Laminin G-like domains, and a short intracellular domain in which the FERM- and PDZ-binding domains are conserved. In contrast, CRB3 is a much smaller family member, with a conserved intracellular domain but a very short extracellular domain (59 amino acids in human) lacking the EGF repeats and Laminin G-like domains. CRB1 and CRB2 are expressed in the retina and brain, with additional expression of CRB2 in the kidney.[16,17] CRB3 is broadly expressed in epithelial tissues and skeletal muscles.[18,19] Human CRB1 is crucial for establishment of cell polarity in retinal cells. Loss of function mutations in CRB1 have been identified as the causative mutations of blindness in patients with retinitis pigmentosa (RP) and Leber congenital

amaurosis (LCA) due to shortening and degeneration of photoreceptors.[16,20–22] Human CRB2 does not seem to be essential for retina development, since mutations in this protein are not a common cause of RP and LCA.[17] In mice, simultaneous depletion of CRB1 and CRB2 causes more severe retinal abnormalities then depletion of either of them alone, indicating at least partial functional redundancy.[23] CRB3 plays a crucial role in tight junction formation and apical membrane morphogenesis.[19,24] Expression of exogenous CRB3 induces the formation of tight junctions at apical cell junctions of human mammary cells, which normally do not form tight junctions.[25] Overexpression of CRB3 in frog blastomeres or in monolayers of MDCK cells causes expansion of the apical domain, while overexpression of CRB3 in MDCK cells in 3D culture leads to disruption of cell polarity.[26,27] Instead of forming a polarized cyst containing a central lumen, MDCK cells overexpressing CRB3 formed aggregates of cells that lacked a lumen and showed disorganized localization of junctional and apical markers. Overexpression of a CRB3 variant lacking the PDZ-binding domain did not result in any abnormalities.[26] Knockdown of CRB3 in MDCK cells grown in 3D culture also led to a no-lumen phenotype, indicating that CRB3 is involved in formation of the first apical membrane at the 2-cell stage, which guides formation of the lumen and encloses the lumen.[28,29] Crb3 knockout mice die shortly after birth from epithelial defects, such as cystic kidneys and abnormal intestine with apical membrane blebs and disrupted microvilli.[30]

In *Caenorhabditis elegans* two Crumbs proteins have been described: CRB-1 and EAT-20. CRB-1 expression starts during embryonic development and the protein localizes to the apical domain of intestinal and pharyngeal cells.[31,32] CRB-1 in the embryonic intestine localizes to a subapical band just apical of the junctional protein DLG-1 and overlapping with the apical protein IFB-2.[32] Loss of *crb-1* does not cause overt defects in polarity. An indication for a more subtle role in cell polarity for CRB-1 comes from studies examining the roles of the *C. elegans* Scribble homolog LET-413 and the *C. elegans* α-catenin homolog HMP-1 in positioning of DLG-1. Depletion of LET-413 results in disrupted positioning of DLG-1, while DLG-1 localization appears normal in *let-413 hmp-1* double knock down embryos. Triple *let-413 hmp-1 crb-1* RNAi leads to a similar phenotype as *let-413* RNAi.[32] Thus, there might be redundancy between the Cadherin-Catenin complex (CCC) and CRB-1. EAT-20, the other *C. elegans* Crumbs protein, is expressed in the pharynx, in a subset of neurons, and in hypodermal cells.[33,34] A presumable null mutant of *eat-20* suffers from reduced pharyngeal pumping. The mutant worms have a starved appearance, a smaller brood size, and a prolonged egg-laying period.[33] *crb-1* and/or *eat-20* RNAi embryos develop normal epithelial identity.[31,32] No essential role in polarity regulation has been uncovered for Crumbs in *C. elegans*.

Here, we further investigate the role of Crumbs proteins in *C. elegans*. We characterize a candidate third Crumbs homolog, which is highly similar to mammalian CRB3 in size and structure. We show that this homolog of Crumbs is expressed in several tissues in the embryo and larval stages and that the protein localizes apical in the intestine and pharynx. Targeted deletion of all three Crumbs homologs with CRISPR/Cas9 did not result in embryonic lethality. Upon overexpression of Crumbs a fraction of PAR-3 mislocalizes cytoplasmic, though the localization of two other cell polarity proteins, DLG-1 and LGL-1, is

unaffected. These results suggest that *C. elegans* Crumbs proteins play a non-essential role in the establishment of epithelial polarity in *C. elegans*.

RESULTS

Identification of a candidate *C. elegans* CRB3 homolog

Two Crumbs homologs have been described in *C. elegans*: *crb-1* and *eat-20*. Of the two corresponding proteins, the CRB-1 protein is most similar in size and protein domain composition to *Drosophila* Crumbs (Fig. 1). CRB-1 consists of 1722 amino acids and contains 26 EGF repeats and two Laminin G-like domains in its extracellular part. The EAT-20 protein is 808 amino acids long and comprises three EGF repeats in its extracellular part. Both these Crumbs proteins have a long extracellular part, a transmembrane domain, and a short intracellular part. The essential residues of the FERM-binding motif and the residues of the PDZ-binding motif in the intracellular domain of the *Drosophila* Crumbs protein are conserved in both CRB-1 and EAT-20.[9,35] Since depletion of CRB-1 and EAT-20 did not result in a severe phenotype and depletion of Crumbs causes lethality in *Drosophila*, we were interested whether there are any additional *C. elegans* Crumbs homologs that might act redundantly with CRB-1 and EAT-20. To identify potential additional Crumbs homologs, we searched the predicted *C. elegans* proteome for candidate homologs of Crumbs proteins by BLAST. A BLASTP search with the human CRB3 sequence yielded in addition to CRB-1 and EAT-20 a third significant hit, C35B8.4. Large-scale expression profiling experiments indicated that the C35B8.4 gene is expressed.[36,37] The predicted protein encoded by C35B8.4 is 100 amino acids long, similar in length to mammalian CRB3, and has a short extracellular tail without recognizable domains, followed by a transmembrane domain and an intracellular part. The tyrosine on position 10 and the glutamic acid on position 16 of the intracellular part, located within the FERM-binding domain, were essential for rescuing Crumbs null phenotypes in *Drosophila* and these residues are conserved in *C. elegans* C35B8.4.[9,35] The final 4 amino acids of CRB-3 are EGLI, and hence differ from the canonical ERLI PDZ-binding motif present in most Crumbs proteins. However, an alternative splice variant of human CRB3 also contains a different C-terminus and was shown to function in spindle assembly, cilia formation, and cell division. This alternative splice variant binds to importin β-1, unlike the ERLI isoform.[38] Thus, the final four amino acids of C35B8.4 could potentially have a different binding specificity. Based on the similarity of C35B8.4 to human CRB3 and the apical localization of the protein described below, we assigned C35B8.4 the name *crb-3*.

CRB-3 localizes apically in multiple polarized tissues

To determine a potential role for CRB-3 in establishing epithelial polarity, we first determined its expression pattern and subcellular localization. If CRB-3 acts as a regulator of epithelial polarity similar to other crumbs proteins and mammalian CRB3, we expected it to localize at the apical membrane domain of epithelial cells. To visualize the expression and localization pattern of CRB-3, we fused GFP to the 3' end of the predicted *crb-3* gene by recombineering, a fosmid-based tagging approach, to maintain most of the endogenous regulatory sequences.[39] We generated a transgenic line carrying an integrated copy of this

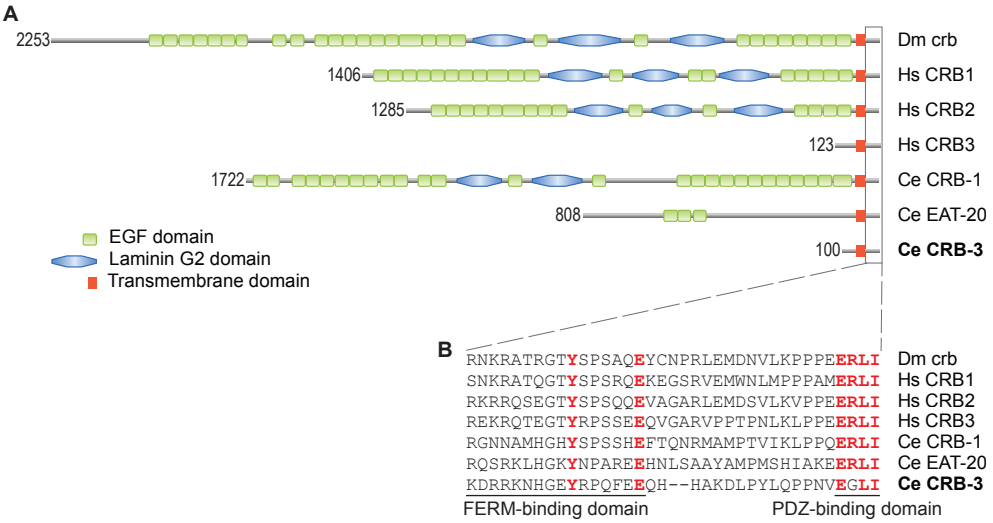


Figure 1. Homology between *Drosophila*, human, and *C. elegans* Crumbs proteins. (A) Protein domain structure of the Crumbs proteins. The number in front of the protein corresponds to the length of the protein in amino acids. (B) Intracellular part with essential residues of the FERM-binding and PDZ-binding domains depicted in red. Dm = *Drosophila melanogaster*, Hs = *Homo sapiens*, Ce = *Caenorhabditis elegans*.

construct by gamma-irradiation mediated integration of an extrachromosomal array. Two independently integrated strains were obtained, showing the same expression pattern. CRB-3::GFP was first detected in embryonic pharyngeal and intestinal precursor cells (Fig. 2). Throughout the larval stages the fusion protein localized to the apical membrane domain of pharyngeal cells, to the excretory canal, to the apical membrane domain of intestinal cells, to a circumferential pattern resembling the pattern of commissural axons, in the dorsal and ventral nerve cords, to the coelomocytes, and frequently (n=4/6) to the apical membrane domain of the rectal epithelium. During the fourth larval stage CRB-3::GFP became visible in the uterus. CRB-3::GFP was not detected in the seam cells, an epithelial tissue in which EAT-20 is known to localize apically. The apical localization of CRB-3::GFP in the pharynx and intestine strengthens our hypothesis that CRB-3 is indeed a *bona fide* Crumbs homolog.

Deletion of all three Crumbs homologs does not cause lethality

Compared to *Drosophila*, where loss of *crumbs* results in complete disruption of polarity in epithelial tissues and embryonic lethality, the phenotypes of *crb-1* and *eat-20* mutants are mild. The partial *eat-20* deletion allele, *nc4*, causes a frame shift and a premature stop. The corresponding truncated protein lacks a long stretch of the extracellular part and the entire transmembrane domain and intracellular part. The strain carrying this mutation, ST6, has a starved appearance with a shorter body length, reduced brood size and an extended egg laying period.[33] The partial deletion allele of *crb-1*, *ok931*, results in an in-frame deletion in the extracellular part. No phenotype has been observed for this mutation. We obtained a partial deletion mutant of *crb-3* from the National Bioresource Project in Japan (*tm6075*). This mutation is predicted to result in a frame shift and premature stop. The corresponding

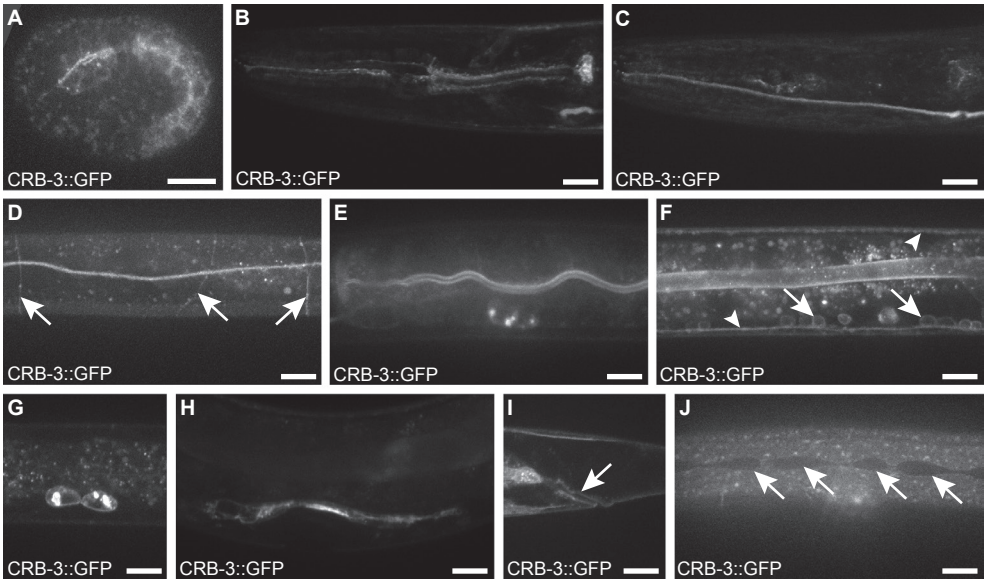


Figure 2. Expression and localization of CRB-3::GFP throughout development. (A) 1.5-fold embryo, (B-G, J) third larval stage, (H, I) fourth larval stage. (B) pharynx, (C) excretory canal, (D) circumferential pattern, indicated by arrows, resembling the pattern of commissural axons, (E) apical localization in intestinal cells (CRB-3::GFP is also visible in the coelomocytes), (F) dorsal cord and ventral nerve cords indicated by arrowheads and cell bodies of the ventral nerve cord motor neurons indicated by arrows, (G) coelomocytes, (H) uterine epithelial cells, (I) rectal epithelium as indicated by the arrow, (J) seam cells indicated by arrows. Scale bars reflect 10 μm.

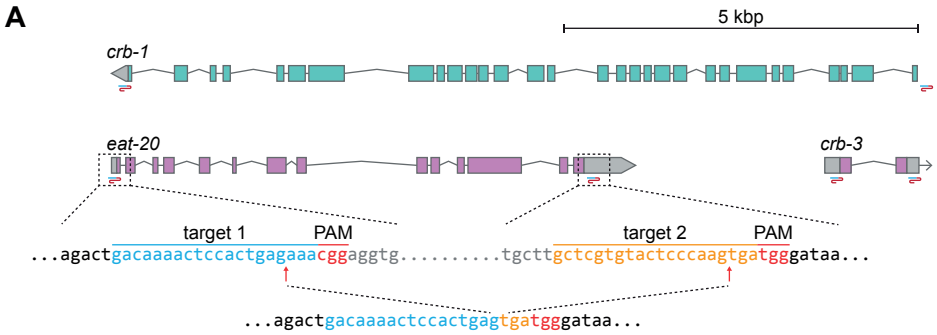
protein lacks the transmembrane domain and intracellular part. We checked for embryonic lethality in this mutant strain and did not observe any. The absence of a strong phenotype upon mutation of any of the three Crumbs homologs could be caused by redundancy between the Crumbs homologs. Therefore, we generated a strain lacking all three genes. No genetic null allele exists for *crb-1*, and traditional methods for generation of deletion mutants are time consuming and cannot precisely delete a chosen sequence. Furthermore, all three Crumbs homologs lie on the X chromosome, relatively close together and near the end of the chromosome, complicating the generation of a triple mutant. For these reasons, we developed a CRISPR/Cas9 approach to rapidly delete entire loci, enabling us to sequentially generate complete loss-of-function alleles for all three Crumbs homologs. CRISPR/Cas9 is a method where a single guide RNA (sgRNA) guides the Cas9 endonuclease to a homologous site in the genome, where Cas9 induces a double strand break (DSB). Previously, we used CRISPR/Cas9 to target a single DSB to specific loci in the genome, which results in the generation of small insertions or deletions due to repair errors during non-homologous end joining.[40] We adapted our CRISPR/Cas9 approach to delete entire loci (Fig. 3). By using two sgRNAs, one targeting a sequence before the start codon of the gene and the other targeting a sequence after the stop codon, the intervening sequence can be lost during DNA repair. Deletions of genes can easily be detected in the F1 generation by PCR with primers

flanking the desired deletion. To generate a triple Crumbs knockout strain, we sequentially deleted the three Crumbs homologs (Fig. 3). We started from the *eat-20(nc4)* background, since this mutation is likely a molecular null. To delete the ~11 kb *crb-1* coding sequence, we injected expression constructs for two different sgRNAs (*U6::sgRNA*), Cas9 controlled by the heat shock promoter (*Phsp-16.48::Cas9*) and a coinjection marker (*Pmyo-3::mCherry*) in the gonad of 30 P0 animals and exposed the injected animals to a 1 hour heat shock at 34°C. We screened 89 transgenic F1 worms for deletion of the gene by PCR and obtained one deletion mutant. DNA sequence analysis confirmed the deletion with boundaries close to the predicted Cas9 cut sites, thereby eliminating the entire *crb-1* coding region (Fig. 3B). The homozygous *eat-20 crb-1* double mutant did not show embryonic lethality. Next, we used this double mutant as a background to delete the *crb-3* coding sequences using the same CRISPR/Cas9 approach. We obtained 3 *crb-3* deletion alleles out of 84 transgenic F1 worms, where the entire coding region was deleted. Sequence analysis confirmed that the deletions had boundaries close to the predicted Cas9 cut sites (Fig. 3B). Again, we observed no embryonic lethality or overt larval defects in this triple Crumbs deletion mutant strain. Finally, to ensure that no functional EAT-20 protein is produced, we deleted the remaining *eat-20* sequences, with a success rate of 13 deletion mutants out of 32 transgenic worms. We sequenced a deletion mutant that appeared to be homozygous based on the absence of the wild type PCR band and again confirmed that the deletion boundaries were close to the predicted Cas9 cut sites (Fig. 3B). The homozygous triple Crumbs deletion mutant did not show any embryonic lethality or overt larval defects. Thus, despite evolutionary conservation, the function of the Crumbs family members is not essential for viability and does not seem to be required for apical-basal polarity in *C. elegans* epithelia.

Overexpression of Crumbs leads to changes in the localization of PAR-3

Overexpression of *Drosophila* Crumbs results in expansion of the apical membrane domain.[5] This phenotype can also be established by overexpression of a fusion construct in *Drosophila* containing the intracellular part of *C. elegans* CRB-1.[9] We wondered whether overexpression of Crumbs family members can also induce polarity defects in *C. elegans*. Therefore, we designed overexpression constructs for all three Crumbs homologs. To be able to choose the moment of overexpression, we placed the Crumbs genes under the control of a heat shock inducible promoter. To detect changes in cell polarity we made use of several transgenic strains with markers of epithelial polarity. We used integrated transgenic strains expressing the junctional marker DLG-1::GFP, the basolateral marker LGL-1::GFP, and the apical marker PAR-3::GFP. We injected a combination of all three Crumbs homologs in each of these marker strains and looked for any phenotype. Based on the fluorescent co-injection marker, we selected strains with a high transmission rate of the extra-chromosomal array carrying the overexpression constructs. After heat shock at the L2 stage, we examined the localization pattern of DLG-1::GFP, LGL-1::GFP, and PAR-3::GFP in the pharynx, the seam cells and the intestine two, four, six and 24 hours after heat shock. Localization of DLG-1::GFP and LGL-1::GFP was not altered at any of the time points in any of the tissues (Fig. S1). However, overexpression of Crumbs did result in a change in PAR-3::GFP localization from the apical site of the pharyngeal cells to the cytoplasm (Fig. 4). We first observed this

phenotype 2 hours after heat shock treatment. Surprisingly, overexpression of Crumbs also led to a cytoplasmic PAR-3::GFP signal in the intestine, even though no GFP signal was detected in this tissue in the parental strain. This could be due to stabilization of PAR-3::GFP in the cytoplasm. In control experiments, we examined the localization of PAR-3::GFP and the other polarity markers after heat shock treatment in the absence of Crumbs overexpression constructs. This never induced changes in the subcellular localization of a polarity marker (Fig. S1 and Fig. 4). The localization of the markers was also never altered in animals carrying the extra-chromosomal array that were not subjected to heat shock treatment (data not shown). The changed localization of PAR-3::GFP upon overexpression of the Crumbs homologs suggests a role for Crumbs in polarity regulation in *C. elegans*.



B

crb-1

tatccgacaatacacctggctctctaggacctt...cctctcctgttcatctgtgtctttttctcgttgc	Target sequences
tatccgacaatctgtgtctttttctcgttgc	Deletion

eat-20

agactgacaaaactccactgagaaacggaggtg...tgcttgctcgtgtactcccaagtgatgggataa	Target sequences
agactgacaaaactaa	Homozygous deletion obtained

crb-3

taccggaaaatggcgtcaaacagtacggttggc...acattgaattagtctcgctttgccttggtcatc	Target sequences
taccggaaaatggcgtcaaacaccttggtcaaaccttggtcatc	Deletions obtained
taccggaaaatggcgtcaaaccttggtcatc	
taccggaaaatggcgtcaaacttggtcatc	

Figure 3. Creation of Crumbs deletions through CRISPR/Cas9. (A) Strategy to delete the entire coding sequences of the three Crumbs genes. Cas9 is targeted to two sequences flanking the coding region of a gene which induces DSBs at the target sites. The intervening sequence may be lost after repair of the DSBs. (B) The nucleotide sequence of the target sites and the sequences after deleting the coding region as confirmed by sequence analysis for each of the Crumbs genes. The PAM sequences are depicted in red, the CRISPR/Cas9 target sequences before the start codon in blue, and the CRISPR/Cas9 target sequence after the stop codon in orange.

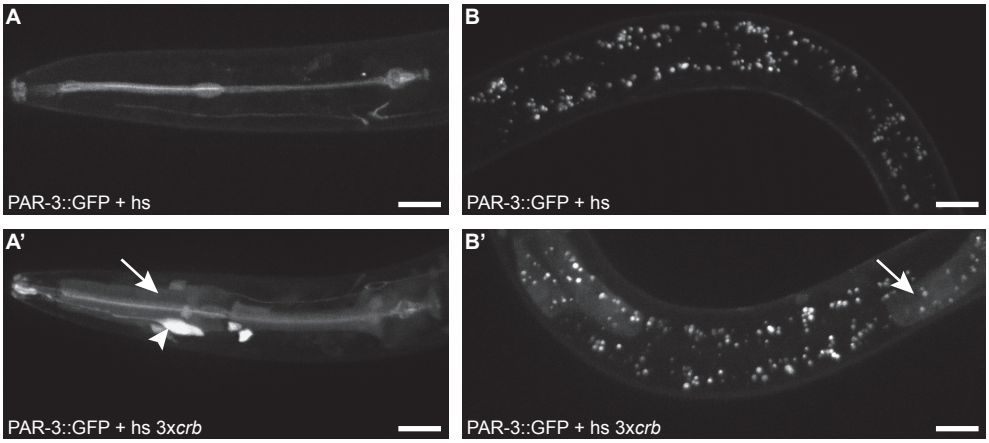


Figure 4. Crumbs overexpression in the pharynx and intestine of a PAR-3::GFP marker strain. (A, B) Controls without overexpression, (A', B') with triple Crumbs overexpression. (A, A') Pharynx of the PAR-3::GFP marker strain. Cytoplasmic PAR-3::GFP localization is indicated with an arrow. The neuronal co-injection marker of the overexpression array is indicated by the arrowhead. (B, B') Intestine of the PAR-3::GFP marker strain. The cytoplasmic PAR-3::GFP is indicated with the arrow. All worms depicted are in the third larval stage. Scale bars reflect 10 μm.

DISCUSSION

In this chapter, we identified a third *C. elegans* Crumbs homolog, which we termed CRB-3, based on similarity to mammalian CRB3. We observed expression of *C. elegans crb-3* in several tissues in the embryo and larval stages, with apical localization of CRB-3::GFP in the intestine and pharynx. All three *C. elegans* Crumbs homologs were deleted using a novel CRISPR/Cas9 approach, but this did not result in lethality of the triple deletion mutant. Overexpression of Crumbs induced PAR-3::GFP localization in the cytoplasm, though DLG-1::GFP and LGL-1::GFP localization was unaffected. These results show that Crumbs has no essential role in *C. elegans*, although the effects of Crumbs overexpression on PAR-3 indicates a more subtle function for Crumbs in cell polarity.

The degree to which Crumbs is essential for viability seems to differ between organisms. Proper Crumbs functioning is indispensable for epithelial polarity establishment and maintenance in *Drosophila* and interfering with the functioning of Crumbs leads to lethality at an early stage when epithelia are formed.[4] In human, mutations in CRB1 cause severe polarity defects in the retina, the major tissue of expression.[16,20–22] CRB2 is also expressed in this tissue and does not seem to be essential for cell polarity regulation.[17] No lethal mutations were reported for Crumbs in human. Crb3 knockout mice die shortly after birth, thus depletion of Crb3 does not result in complete obstruction of epithelial polarity in this model system, though Crb3 is the main expressed Crumbs homolog in epithelia.[30] Our results suggest that the *C. elegans* Crumbs homologs have an even less essential function, since deletion of all three Crumbs homologs did not cause lethality.

The lack of a critical function for the three *C. elegans* Crumbs homologs points to redundant mechanisms that probably are active during the regulation of epithelial cell polarity. A likely candidate for such a redundant mechanism is the apical PAR complex. In future experiments we will therefore examine whether the triple Crumbs deletion predisposes epithelial cells to loss of polarity upon inactivation of *par-3*, *par-6*, or *pkc-3*. Other candidates are the members of the CCC complex, *hmp-1*, *hmp-2*, and *hmr-1*, since it was shown that CRB-1 might act redundantly with this complex in positioning DLG-1.[32] It is also possible that deletion of the Crumbs genes leads to a subtle phenotype in *C. elegans*. This can be investigated by crossing the triple deletion mutant with transgenic strains expressing fluorescently labeled cell polarity proteins and examining whether there are any changes in the localization of these markers in the absence of the Crumbs proteins.

PAR-3::GFP was present in the cytoplasm upon overexpression of Crumbs (Fig. 4), which could be explained by an excess of Crumbs protein competing PAR-3::GFP off the cortical PAR-6/PKC-3 complex. In *Drosophila*, polarization of epithelia is a stepwise process. During the initial phase of epithelial polarity establishment Baz, the *Drosophila* PAR-3 homolog, recruits aPKC, PAR-6, and the Crumbs complex to the plasma membrane.[41–44] Later, Baz is excluded from the apical domain in epithelial tissues. This depends on two exclusion mechanisms: phosphorylation of Baz by aPKC disrupts the interaction between these two proteins, and the binding of the Crumbs complex to PAR-6 prevents the Baz/PAR-6 interaction.[45–48] In the absence of Crumbs or aPKC phosphorylation of Baz, mislocalized Baz recruits adherens junction components apically, leading to a loss of the apical domain and an expansion of the lateral domain.[41,45] Thus, apical exclusion of Baz by Crumbs and aPKC defines the apical/lateral border. In *C. elegans* PAR-3 exclusion from the PAR-6 apical domain is also observed in embryonic intestinal cells and cells of the developing vulva, where PAR-3 localizes subapical.[49,50] It would be interesting to investigate the changes in localization of Crumbs upon overexpression of PAR-3. If there is indeed competition between the Crumbs homologs and PAR-3 for binding to PAR-6, an excess of PAR-3 could alter the localization of Crumbs. To test this, we will inject a heat shock inducible overexpression construct of PAR-3 into a CRB-3::GFP strain and examine the localization pattern of CRB-3::GFP upon overexpression of PAR-3. Another important question is whether the PAR-3::GFP mislocalization phenotype can be achieved by overexpression of a single Crumbs homolog or if it requires overexpression of a certain combination of Crumbs homologs. Generating and examining strains with all possible double combinations and the single Crumbs overexpression constructs in the PAR-3::GFP background will answer this question.

The other core members of the Crumbs complex, Stardust (Sdt), PATJ, and Lin7, are conserved to some degree in *C. elegans*. In *C. elegans* there are three candidate Sdt homologs: *magu-1*, *magu-2*, and *magu-3*.[51,52] A mutant is only available for MAGU-2, but this probable null mutant does not exhibit any obvious phenotype. Depletion of *Drosophila* Sdt causes similar polarity defects in epithelia as are caused by the depletion of Crumbs.[4,44,48] PATJ and Lin7 both have one homolog in *C. elegans*, MPZ-1 and LIN-7, respectively.[51,52] Mutations in both corresponding genes can cause lethality, although the arrest stages and precise phenotypes were not determined. LIN-7 localizes to the apical junctions of the vulva

epithelium.[53] Absence of Lin7 does not result in epithelial polarity defects in *Drosophila*.[54] PATJ is only necessary for cell polarity regulation in *Drosophila* photoreceptor cells.[55] No relation to Crumbs was described for any of the *C. elegans* Sdt/PATJ/Lin7 homologs.

Although the Crumbs, PAR, and Scribble groups of polarity regulators are active in a variety of cell types, the mechanisms through which they establish polarity appears to vary markedly in different cell types or situations. For example, in *Drosophila*, not all epithelia in which Crb is expressed require Crb to maintain epithelial polarity.[56] Similarly, at least three groups of basolateral regulators function at different times during embryonic development of *Drosophila*.[56] Our studies of the functioning of the Crumbs complex in *C. elegans* should contribute to our understanding how cortical polarity regulators establish polarity in different situations.

MATERIALS AND METHODS

Protein domain prediction

Interpro was used as a tool for prediction of protein domains to be able to compare Crumbs homologs on a level of protein domain composition.[57] For *eat-20* splice variant a was used for protein domain prediction.

Generation of GFP fusion constructs

The CRB-3::GFP construct was generated according to the recombineering procedure published before.[39] We amplified GFP with affinity tag using the primers SW_rc_coc_C35B8.4_F (acgcaaaagacctaccatatcttcaacctccgaatgtagaaggacttatcggagggatctgaggaggatctggaggagga) and SW_rc_coc_C35B8.4_R (cacatataaaagcgcccaatttgattgaaatgaataataaaaatattttattcatgccattcaatcttctgagcttcg) with KOD hot start polymerase. The annealing temperature used was 70°C. The PCR product was recombined into fosmid WRM0628dH07. The resulting CRB-3::GFP is available on request.

Scoring embryonic lethality

Individual adult animals were allowed to lay eggs overnight at 20°C. The next day the adults were taken off and eggs were let to hatch at 20°C. The next day the plates were checked for unhatched embryos.

CRISPR/Cas9

To generate deletion alleles of *crb-1*, *eat-20*, and *crb-3*, we simultaneously targeted a site near the start codon and a site near the stop codon of each gene with CRISPR/Cas9. To clone the sequences of the target sites into the sgRNA expression vector, we first annealed pairs of oligonucleotides crb-1_CRISPR_1_F (aattgacaatacacctggctctct) with crb-1_CRISPR_1_R (aaacagagagccaggtgtattgtc), crb-1_CRISPR_2_F (aattgagaaaagacacagatgaac) with crb-1_CRISPR_2_R (aaacgttcatctgtgtcttttctc), eat-20_CRISPR_1_F (aattgacaaaactccactgagaaa) with eat-20_CRISPR_1_R (aaactttctcagtggagttttgtc), eat-20_CRISPR_2_F (aattgctcgtgtactcccaagtga) with eat-20_CRISPR_2_R (aaactcacttgggagtacacgagc), crb-3_CRISPR_1_F (aattgaaaatggcgtcaaacagta) with crb-3_CRISPR_1_R (aaactactgtttgacgccattttc), and crb-3_CRISPR_2_F (aattgaattagtctcgctttgcct) with crb-3_CRISPR_2_R (aaacaggcaaagcgagactaattc). The resulting linkers were ligated into the *Bsa*I digested U6::sgRNA expression vector pMB70. For each deletion, we injected 30 animals with a mixture containing 5 ng/µl *Pmyo-3::mCherry* (pCFJ104, Addgene #19328), 50 ng/µl of each of the two sgRNAs, and 50 ng/µl *Phsp-16.48::Cas9* using standard *C. elegans* microinjection procedures.

To induce expression from the *hsp-16.48* promoter, injected animals were heat shocked for 1 hour at 34°C on agar plates floating in a water bath, 30 min after injection. From transgenic F1 animals expressing mCherry, we PCR amplified a region surrounding the target site using primers crb-1_CRISPR_check_F (gtcgcttgttatgggataaaac) and crb-1_CRISPR_R (ggtaccagtgacaacatttgct) for *crb-1*, eat-20_CRISPR_check_F (gtgtgaccaaacttattgctttc) and eat-20_CRISPR_check_R (gctctccaagtcaaaaagttctta) for *eat-20*, and crb-3_CRISPR_check_F (ggagacggagatggtcaagt) and crb-3_CRISPR_check_R (acgtgtagtactcggtgttcagg) for *crb-3*. Homozygous mutant lines were isolated from single F2 animals and their genotype was determined by PCR and sequence analysis.

Crumbs heat shock constructs

The genomic sequences of the *crb-1*, *eat-20*, and *crb-3*, together with the heat shock promoter *Phsp-16.41* and the *tbb-2* 3' UTR were cloned into the pBSK vector using standard restriction enzyme cloning. The inserts were obtained by PCR with KOD hot start polymerase. The specific restriction sites were added to the primers. *eat-20* and *crb-3* sequences were inserted in one part using primers eat-20_BamHI_F (aaggatccaaaaatgaccacgttttgtcgagt), eat-20_SacI_R (aagagctcttagatcagccgctcctcct), crb-3_BamHI_F (aaggatccaaaaatggcgtcaaacagtacggt) and crb-3_SacI_R (aagagctcttagataagtccttctacattcgga). The *crb-1* gene was inserted in three parts in subsequent cloning rounds. For *crb-1* we used primers crb-1_BamHI_F1 (aaggatccaaaaatgaaatatcaacattttctcatatttt)andcrb-1_MscI_R1(aatggccacttttgcacacctgtgtcttg)forthefirstcloninground, crb-1_MscI_F2 (aatggccaatgaggacacgcggagta) and crb-1_NheI_R2 (aagctagcagtatcgacctcgccccta) for the second cloning round, and crb-1_NheI_F3 (aagctagcgtagctcgcgcattttgat) and crb-1_NcoI_R3 (aaccatggtcagataagacgttcttgaggtg) for the final cloning round. The final sequences of these constructs are available on request.

Culture conditions and strains

C. elegans strains were maintained under standard culture conditions as previously described.[58] Plasmids were injected and gamma–integration was performed using standard *C. elegans* procedures. The following strains were used:

ST6: *eat-20(nc4)*X

RB1011: *crb-1(ok931)*X

BOX143: *crb-3(tm6075)*X

BOX41: *mibIs23[lgl-1::GFP, Pmyo-3::mCherry]*V

BOX42: *mibIs24[crb-3::GFP, Pmyo-3::mCherry]*IV

BOX56: *mibIs31[dlg-1::GFP, Pmyo-3::mCherry]*V

BOX66: *mibIs41[crb-3::GFP, Pmyo-3::mCherry]*III

BOX80: *mibIs42[par-3::GFP, Pmyo-3::mCherry]*IV

BOX82: *mibIs26[par-3::GFP, Pmyo-3::mCherry]*IV, *mibEx6[Phsp-16.41::crb-1, Phsp-16.41::eat-20, Phsp-16.41::crb-3, Plin-48::GFP]*

BOX83: *mibIs26[par-3::GFP, Pmyo-3::mCherry]*IV, *mibEx7[Phsp-16.41::crb-1, Phsp-16.41::eat-20, Phsp-16.41::crb-3, Plin-48::GFP]*

BOX84: *mibIs26[par-3::GFP, Pmyo-3::mCherry]*IV, *mibEx8[Phsp-16.41::crb-1, Phsp-16.41::eat-20, Phsp-16.41::crb-3, Plin-48::GFP]*

BOX85: *mibIs23[lgl-1::GFP, Pmyo-3::mCherry]*V, *mibEx9[Phsp-16.41::crb-1, Phsp-16.41::eat-20, Phsp-16.41::crb-3, Plin-48::GFP]*

BOX86: *mibIs23[lgl-1::GFP, Pmyo-3::mCherry]*V, *mibEx10[Phsp-16.41::crb-1, Phsp-16.41::eat-20, Phsp-16.41::crb-3, Plin-48::GFP]*

BOX89: *mibIs31[dlg-1::GFP, Pmyo-3::mCherry]*V, *mibEx13[Phsp-16.41::crb-1, Phsp-16.41::eat-20, Phsp-16.41::crb-3, Plin-48::mCherry]*

BOX90: *mibIs31[dlg-1::GFP, Pmyo-3::mCherry]*V, *mibEx14[Phsp-16.41::crb-1, Phsp-16.41::eat-20, Phsp-16.41::crb-3, Plin-48::mCherry]*

BOX142: *crb-1(mib3), eat-20(mib5), crb-3(mib4)*X

Heat shock induction of Crumbs overexpression constructs

Individual adult animals were allowed to lay eggs overnight at 20°C. The next day the adults were taken off and eggs were let to hatch at 20°C. The next day L1 and L2 larvae were heat shocked by incubating sealed plates in a water bath of 34°C for 1 hour. After heat shock, the plates were returned to 20°C.

Microscopy and image processing

Imaging was performed on a spinning disc confocal sytem, consisting of a Nikon Ti-U inverted microscope with a motorized stage and a Piezo Z stage, and a PLAN APO VC 60X oil objective; a Yokogawa CSU-X1 spinning disk unit, equipped with a dual dichroic mirror set for laser wavelengths 488 nm and 561 nm; 488 nm and 561 nm solid state 50 mW lasers, controlled by an Andor revolution 500 series AOTF Laser modulator and combiner; Semrock 525 and 617 nm single band fluorescence emission filters (30 and 73 nm bandwith respectively); Semrock 525 single band fluorescence filter (center wavelength of 525 nm, with a GMBW of 30 nm); and an Andor iXON DU-885 monochrome EMCCD+ camera. All imaging was done using Andor iQ imaging software version 1.1. Maximum projections were generated from a series of slices of a Z-stack with ImageJ and processed with Adobe Photoshop CS6 and Adobe Illustrator CS6.

ACKNOWLEDGEMENTS

We thank Sander van den Heuvel and Adri Thomas for critically reading this manuscript. Some strains were provided by the CGC, which is funded by NIH Office of Research Infrastructure Programs (P40 OD010440). The *crb-3(tm6075)* mutant allele was provided by the MITANI Lab through the National BioResource Project of the MEXT, Japan. This work was supported by Innovational Research Incentives Scheme Vidi grant 864.09.008, financed by the Netherlands Organization for Scientific Research (NWO/ALW).

REFERENCES

1. St Johnston, D. & Ahringer, J. Cell polarity in eggs and epithelia: parallels and diversity. Cell 141, 757–74 (2010).
2. Jürgens, G., Wieschaus, E., Nüsslein-Volhard, C. & Kluding, H. Mutations affecting the pattern of the larval cuticle in *Drosophila* melanogaster II. Zygotic loci on the third chromosome. Dev. Biol. 193, 283–95 (1984).
3. Tepass, U., Theres, C. & Knust, E. crumbs encodes an EGF-like protein expressed on apical membranes of *Drosophila* epithelial cells and required for organization of epithelia. Cell 61, 787–99 (1990).
4. Knust, E., Tepass, U. & Wodarz, A. crumbs and stardust, two genes of *Drosophila* required for the development of epithelial cell polarity. Development 261–8 (1993).
5. Wodarz, A., Hinz, U., Engelbert, M. & Knust, E. Expression of crumbs confers apical character on plasma membrane domains of ectodermal epithelia of *Drosophila*. Cell 82, 67–76 (1995).

6. Izaddoost, S., Nam, S.-C., Bhat, M. A., Bellen, H. J. & Choi, K.-W. *Drosophila* Crumbs is a positional cue in photoreceptor adherens junctions and rhabdomeres. Nature 416, 178–83 (2002).
7. Pellikka, M. *et al*. Crumbs, the *Drosophila* homologue of human CRB1/RP12, is essential for photoreceptor morphogenesis. Nature 416, 143–9 (2002).
8. Knust, E. *et al*. EGF homologous sequences encoded in the genome of *Drosophila melanogaster*, and their relation to neurogenic genes. EMBO J. 6, 761–6 (1987).
9. Klebes, A. & Knust, E. A conserved motif in Crumbs is required for E-cadherin localisation and zonula adherens formation in *Drosophila*. Curr. Biol. 10, 76–85 (2000).
10. Bachmann, A., Schneider, M., Theilenberg, E., Grawe, F. & Knust, E. *Drosophila* Stardust is a partner of Crumbs in the control of epithelial cell polarity. Nature 414, 638–43 (2001).
11. Hong, Y., Stronach, B., Perrimon, N., Jan, L. Y. & Jan, Y. N. *Drosophila* Stardust interacts with Crumbs to control polarity of epithelia but not neuroblasts. Nature 414, 634–8 (2001).
12. Roh, M. H. *et al*. The Maguk protein, Pals1, functions as an adapter, linking mammalian homologues of Crumbs and Discs Lost. J. Cell Biol. 157, 161–72 (2002).
13. Bachmann, A. *et al*. Cell type-specific recruitment of *Drosophila* Lin-7 to distinct MAGUK-based protein complexes defines novel roles for Sdt and Dlg-S97. J. Cell Sci. 117, 1899–909 (2004).
14. Bulgakova, N. A., Kempkens, O. & Knust, E. Multiple domains of Stardust differentially mediate localisation of the Crumbs-Stardust complex during photoreceptor development in *Drosophila*. J. Cell Sci. 121, 2018–26 (2008).
15. Bulgakova, N. A. & Knust, E. The Crumbs complex: from epithelial-cell polarity to retinal degeneration. J. Cell Sci. 122, 2587–96 (2009).
16. Den Hollander, A. I. *et al*. Mutations in a human homologue of *Drosophila* crumbs cause retinitis pigmentosa (RP12). Nat. Genet. 23, 217–21 (1999).
17. Van den Hurk, J. A. J. M. *et al*. Characterization of the Crumbs homolog 2 (CRB2) gene and analysis of its role in retinitis pigmentosa and Leber congenital amaurosis. Mol. Vis. 11, 263–73 (2005).
18. Makarova, O., Roh, M. H., Liu, C.-J., Laurinec, S. & Margolis, B. Mammalian Crumbs3 is a small transmembrane protein linked to protein associated with Lin-7 (Pals1). Gene 302, 21–9 (2003).
19. Lemmers, C. *et al*. CRB3 binds directly to Par6 and regulates the morphogenesis of the tight junctions in mammalian epithelial cells. 15, 1324–33 (2004).
20. Den Hollander, A. I. *et al*. Leber congenital amaurosis and retinitis pigmentosa with Coats-like exudative vasculopathy are associated with mutations in the crumbs homologue 1 (CRB1) gene. Am. J. Hum. Genet. 69, 198–203 (2001).
21. Lotery, A. *et al*. Mutations in the CRB1 gene cause Leber congenital amaurosis. Arch Ophthalmol 119, 415–20 (2001).
22. Mehalow, A. K. *et al*. CRB1 is essential for external limiting membrane integrity and photoreceptor morphogenesis in the mammalian retina. Hum. Mol. Genet. 12, 2179–89 (2003).
23. Pellissier, L. P. *et al*. Targeted ablation of crb1 and crb2 in retinal progenitor cells mimics leber congenital amaurosis. PLoS Genet. 9, e1003976 (2013).
24. Hurd, T. W., Gao, L., Roh, M. H., Macara, I. G. & Margolis, B. Direct interaction of two polarity complexes implicated in epithelial tight junction assembly. Nat. Cell Biol. 5, 137–42 (2003).
25. Fogg, V. C., Liu, C.-J. & Margolis, B. Multiple regions of Crumbs3 are required for tight junction formation in MCF10A cells. J. Cell Sci. 118, 2859–69 (2005).
26. Roh, M. H., Fan, S., Liu, C.-J. & Margolis, B. The Crumbs3-Pals1 complex participates in the establishment of polarity in mammalian epithelial cells. J. Cell Sci. 116, 2895–906 (2003).
27. Chalmers, A. D. *et al*. aPKC, Crumbs3 and Lgl2 control apicobasal polarity in early vertebrate development. Development 132, 977–86 (2005).
28. Torkko, J. M., Manninen, A., Schuck, S. & Simons, K. Depletion of apical transport proteins perturbs epithelial cyst formation and ciliogenesis. J. Cell Sci. 121, 1193–203 (2008).
29. Schlüter, M. A. *et al*. Trafficking of Crumbs3 during cytokinesis is crucial for lumen formation. 20, 4652–63 (2009).
30. Whiteman, E. L. *et al*. Crumbs3 is essential for proper epithelial development and viability. Mol. Cell. Biol. 34, 43–56 (2014).

31. Bossinger, O., Klebes, A., Segbert, C., Theres, C. & Knust, E. Zonula adherens formation in *Caenorhabditis elegans* requires *dlg-1*, the homologue of the *Drosophila* gene discs large. Dev. Biol. 230, 29–42 (2001).
32. Segbert, C., Johnson, K., Theres, C., van Fürden, D. & Bossinger, O. Molecular and functional analysis of apical junction formation in the gut epithelium of *Caenorhabditis elegans*. Dev. Biol. 266, 17–26 (2004).
33. Shibata, Y., Fujii, T., Dent, J. a, Fujisawa, H. & Takagi, S. EAT-20, a novel transmembrane protein with EGF motifs, is required for efficient feeding in *Caenorhabditis elegans*. Genetics 154, 635–46 (2000).
34. Achilleos, A., Wehman, A. M. & Nance, J. PAR-3 mediates the initial clustering and apical localization of junction and polarity proteins during *C . elegans* intestinal epithelial cell polarization. Development 137, 1833–42 (2010).
35. Klose, S., Flores-Benitez, D., Riedel, F. & Knust, E. Fosmid-based structure-function analysis reveals functionally distinct domains in the cytoplasmic domain of *Drosophila* crumbs. G3 (Bethesda). 3, 153–65 (2013).
36. Spencer, W. C. *et al*. A spatial and temporal map of *C. elegans* gene expression. Genome Res. 21, 325–41 (2011).
37. Levin, M., Hashimshony, T., Wagner, F. & Yanai, I. Developmental milestones punctuate gene expression in the *Caenorhabditis* embryo. Dev. Cell 22, 1101–8 (2012).
38. Fan, S. *et al*. A novel Crumbs3 isoform regulates cell division and ciliogenesis via importin beta interactions. J. Cell Biol. 178, 387–98 (2007).
39. Tursun, B., Cochella, L., Carrera, I. & Hobert, O. A toolkit and robust pipeline for the generation of fosmid-based reporter genes in *C. elegans*. PLoS One 4, e4625 (2009).
40. Waaijers, S. *et al*. CRISPR/Cas9-targeted mutagenesis in *Caenorhabditis elegans*. Genetics 195, 1187–91 (2013).
41. Krahn, M. P., Bückers, J., Kastrup, L. & Wodarz, A. Formation of a Bazooka-Stardust complex is essential for plasma membrane polarity in epithelia. J. Cell Biol. 190, 751–60 (2010).
42. McKinley, R. F. A., Yu, C. G. & Harris, T. J. C. Assembly of Bazooka polarity landmarks through a multifaceted membrane-association mechanism. J. Cell Sci. 125, 1177–90 (2012).
43. Franz, A. & Riechmann, V. Stepwise polarisation of the *Drosophila* follicular epithelium. Dev. Biol. 338, 136–47 (2009).
44. Hong, Y., Ackerman, L., Jan, L. Y. & Jan, Y.-N. Distinct roles of Bazooka and Stardust in the specification of *Drosophila* photoreceptor membrane architecture. Proc. Natl. Acad. Sci. U. S. A. 100, 12712–7 (2003).
45. Morais-de-Sá, E., Mirouse, V. & St Johnston, D. aPKC phosphorylation of Bazooka defines the apical/lateral border in *Drosophila* epithelial cells. Cell 141, 509–23 (2010).
46. Harris, T. J. C. & Peifer, M. The positioning and segregation of apical cues during epithelial polarity establishment in *Drosophila*. J. Cell Biol. 170, 813–23 (2005).
47. Walther, R. F. & Pichaud, F. Crumbs/DaPKC-dependent apical exclusion of Bazooka promotes photoreceptor polarity remodeling. Curr. Biol. 20, 1065–74 (2010).
48. Nam, S.-C. & Choi, K.-W. Interaction of Par-6 and Crumbs complexes is essential for photoreceptor morphogenesis in *Drosophila*. Development 130, 4363–72 (2003).
49. Totong, R., Achilleos, A. & Nance, J. PAR-6 is required for junction formation but not apicobasal polarization in *C. elegans* embryonic epithelial cells. Development 134, 1259–68 (2007).
50. Li, B., Kim, H., Beers, M. & Kemphues, K. Different domains of *C. elegans* PAR-3 are required at different times in development. Dev. Biol. 344, 745–57 (2010).
51. Knust, E. & Bossinger, O. Composition and formation of intercellular junctions in epithelial cells. Science 298, 1955–9 (2002).
52. Assémat, E., Bazellières, E., Pallesi-Pocachard, E., Le Bivic, A. & Massey-Harroche, D. Polarity complex proteins. Biochim. Biophys. Acta 1778, 614–30 (2008).
53. Simske, J. S., Kaech, S. M., Harp, S. A. & Kim, S. K. LET-23 receptor localization by the cell junction protein LIN-7 during *C. elegans* vulval induction. Cell 85, 195–204 (1996).

54. Bachmann, A., Grawe, F., Johnson, K. & Knust, E. *Drosophila* Lin-7 is a component of the Crumbs complex in epithelia and photoreceptor cells and prevents light-induced retinal degeneration. Eur. J. Cell Biol. 87, 123–36 (2008).
55. Nam, S.-C. & Choi, K.-W. Domain-specific early and late function of Dpatj in *Drosophila* photoreceptor cells. Dev. Dyn. 235, 1501–7 (2006).
56. Tepass, U. The apical polarity protein network in *Drosophila* epithelial cells: regulation of polarity, junctions, morphogenesis, cell growth, and survival. Annu. Rev. Cell Dev. Biol. 28, 655–85 (2012).
57. Hunter, S. *et al*. InterPro in 2011: new developments in the family and domain prediction database. Nucleic Acids Res. 40, D306–12 (2012).
58. Brenner, S. The genetics of *Caenorhabditis elegans*. Genetics 77, 71–94 (1974).

SUPPORTING INFORMATION

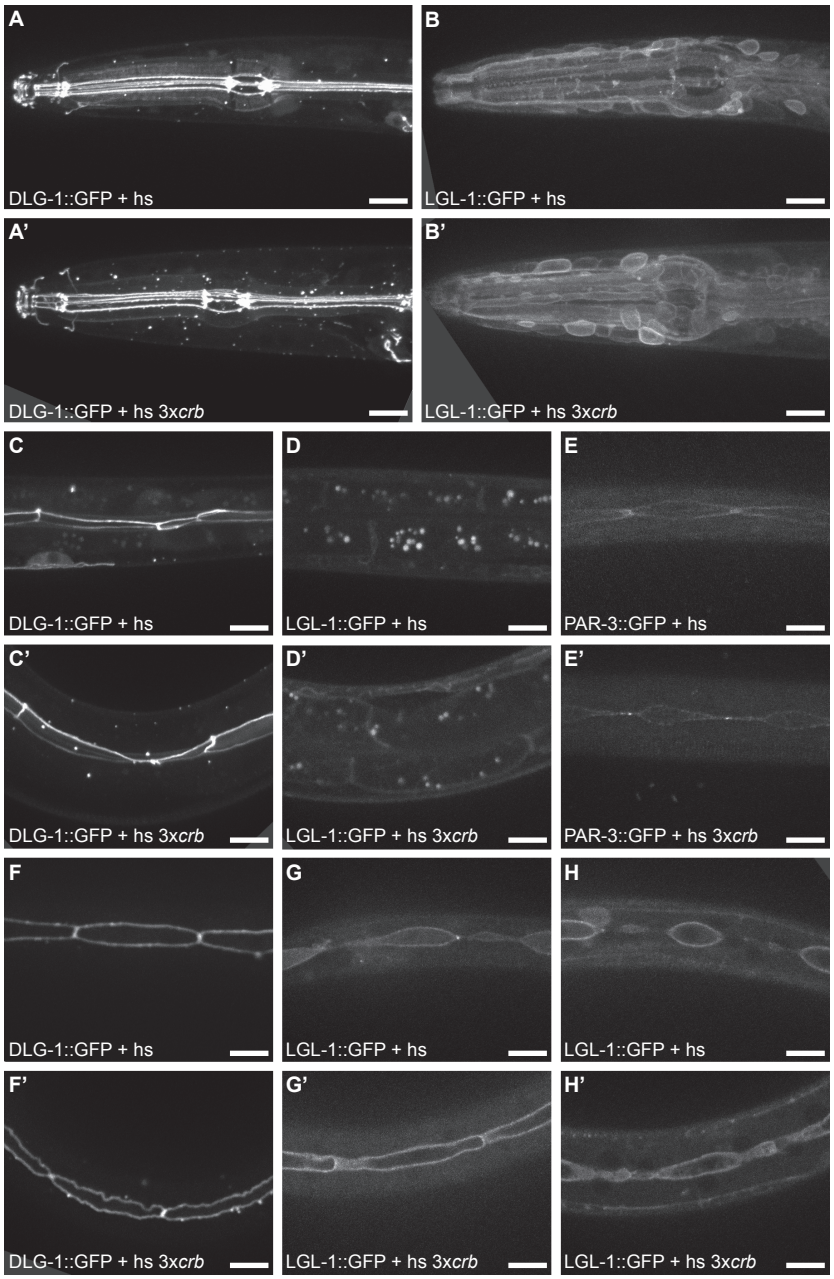


Figure S1. Crumbs overexpression in tissues without phenotype in DLG-1::GFP, LGL-1::GFP, and PAR-3::GFP marker strains. (A, A') Pharynx DLG-1::GFP. (B, B') Pharynx LGL-1::GFP. (C, C') Intestine DLG-1::GFP. (D, D') Intestine LGL-1::GFP. (E, E') Seam cells PAR-3::GFP. (F, F') Seam cells DLG-1::GFP. (G, G') Seam cells LGL-1::GFP, section just below the junctions. (H, H') Seam cells LGL-1::GFP, section at the height of the nuclei. (A, B, C, D, E, F, G, H) without overexpression, (A', B', C', D', E', F', G', H') with triple Crumbs overexpression. All worms depicted are in the third larval stage. Scale bars reflect 10 µm.

Addendum

SUMMARY

Protein-proteininteractions & genome editing

Novel strategies to study cell polarity

Cell polarity is a fundamental property of cells and numerous cellular processes are based on it, from asymmetric cell division to the proper functioning of epithelia and neurons. A number of conserved regulators of cell polarity have been identified, including the PAR, Crumbs, and Scribble groups of cortical polarity regulators. The identification of a core polarity establishment program leads to a number of questions, particularly regarding polarity establishment in multicellular organisms. How are the same components used and integrated in cell type-specific ways to give rise to the wide variety of polarized cell types? What are the downstream components with which the polarity program interacts to establish the final polarized character of cells, including polarized positioning of organelles, functional specification of membrane domains, and polarization of the cytoskeleton? To gain insight into these questions, we developed a tissue-specific affinity purification/mass spectrometry (AP/MS) approach for *C. elegans* based on the *in vivo* biotinylation of Avi-tagged proteins of interest by the bacterial biotin ligase BirA (**Chapter 1**). Tissue-specific biotinylation is accomplished by expressing BirA from a tissue-specific promoter, while the Avi-tagged protein is expressed from its native regulatory sequences. Biotinylated bait proteins are subsequently purified with streptavidin-coated beads and interacting proteins can then be identified by mass spectrometry. We confirmed the tissue-specificity of the biotinylation and purification and applied our approach on several polarity proteins. Purification of DLG-1 from intestine and CDC-42 from the seam cells resulted in the identification of several known interaction partners, demonstrating that our approach can identify valid interactions from specific cells or tissues.

In addition to identifying the components that mediate polarity establishment in different tissues, it is important to gain a detailed mechanistic understanding of the proteins that regulate polarity. An important step toward understanding the functioning of a protein is an analysis of its structure: which regions of a protein mediate which potential functions. The functioning of a protein can in large part be dictated by the interactions it engages in with other proteins. Here, we expand the concept of Y2H-based interaction domain mapping to the human genome-wide level (**Chapter 2**). We generated a human prey library by fragmenting an ORFeome collection with ultrasonication and demonstrated the quality of the library by screening it with polarity and cell division proteins. We identified several interactions previously described in literature as well as novel interactions, and validated 55% of all identified interactions by affinity purifications in cell culture.

A key step after identifying regions or residues of a protein that are predicted to mediate a particular function is to test these predictions *in vivo* in the context of a developing organism. The best approach for such tests is to precisely modify the genome of the organism under study to express a mutated protein that for example lacks potentially important phosphorylation sites or interaction domains. We developed the CRISPR/Cas9 system for *C. elegans* to be able to specifically engineer the genome of *C. elegans* (**Chapter 3**). By expressing a guide RNA the Cas9 endonuclease is escorted to a specific locus in the genome

where it creates a double strand break. Imprecise repair of the break can yield mutations and we obtained mutants for all four genes tested. Several other groups independently adapted the CRISPR/Cas9 system for *C. elegans*, using different approaches that vary in the methods of delivering Cas9 and the guide RNA. We reviewed the different approaches taken for CRISPR/Cas9 in *C. elegans* in **Chapter 4**.

We applied our CRISPR/Cas9 approach to gain a better understanding of the functioning of the Crumbs complex in *C. elegans* (**Chapter 5**). While the Crumbs protein is essential for epithelial polarity and viability in *Drosophila*, the two Crumbs homologs identified in *C. elegans* thus far do not appear to play a major role in polarity establishment. We identified a third Crumbs family member in *C. elegans*, which localizes in a polarized pattern in several tissues. To identify the function of the three Crumbs family members in *C. elegans* development, we developed a variation of our CRISPR/Cas9 approach to delete entire coding regions of genes, and generated a triple Crumbs deletion mutant by sequentially removing the entire coding sequence for each gene. Remarkably, animals lacking all three Crumbs homologs are viable. Overexpression of all three *C. elegans* Crumbs homologs caused changes in the localization pattern of the polarity protein PAR-3, suggesting that the *C. elegans* Crumbs family members play a non-essential role in polarity establishment.

We have developed a number of important tools to study protein function, which have already yielded novel insights into polarity regulation in *C. elegans*, and are key enabling technologies to continuing this line of research in the future.

CURRICULUM VITAE

Selma Waaijers was born on March 11th 1985 in Hilversum, The Netherlands. She attended the 'gemeentelijk gymnasium Hilversum' in 1997 where she passed her VWO exam in 2003. In the same year she started her Bachelor studies Biology at Utrecht University, following courses focused on genetics and writing a Bachelor thesis on asymmetric cell division in the research group of prof. dr. B. Scheres. During her biomedical Master 'Cancer, Genomics, and Developmental Biology' she completed her first rotation project on Tdrd9, a RNA helicase required for germ cell maintenance in adult male zebrafish in the research group of prof. dr. René F. Ketting at the Hubrecht Institute. In 2008 she performed a second rotation project on mapping protein-protein interactions with yeast two-hybrid on a genome wide scale in the research group of prof. dr. Marc Vidal at the Dana Farber Cancer Institute in Boston, USA. She did a final rotation project and wrote her Master thesis on expression profiling using microarrays of DAmP mutants of essential kinases and phosphatases in *Saccharomyces cerevisiae* in the research group of prof. dr. Frank Holstege at the UMCU. In April 2009 she started working in the research group of dr. Mike Boxem on the work described in this thesis.

LIST OF PUBLICATIONS

S. Waaijers & M. Boxem. Engineering the *Caenorhabditis elegans* genome with CRISPR/Cas9. Methods (2014).

B. de Albuquerque, M.J. Luteijn, R.J. Cordeiro Rodrigues, P. van Bergeijk, S. Waaijers, L.J.T. Kaaij, H. Klein, M. Boxem & R.F. Ketting. PID-1 is a novel factor that operates during 21U RNA biogenesis in *Caenorhabditis elegans*. Genes & Development (2014).

S. Waaijers, V. Portegijs, J. Kerver, B.B.L.G. Lemmens, M. Tijsterman & M. Boxem. CRISPR/Cas9-targeted mutagenesis in *Caenorhabditis elegans*. Genetics (2013).

S. Waaijers[1], T. Koorman[1], J. Kerver & M. Boxem. Identification of human protein interaction domains using an ORFeome-based yeast two-hybrid fragment library. Journal of Proteome Research (2013).

Arabidopsis Interactome Mapping Consortium. Evidence for network evolution in an *Arabidopsis* interactome map. Science (2011).

In preparation

S. Waaijers, E. Kruse & M. Boxem. The conserved *C. elegans* Crumbs protein family plays a nonessential role in epithelial polarity.

S. Waaijers, S. S. Goerdayal, S. van den Heuvel, B. Tursun, J. Muñoz, A. J. Heck & M. Boxem. Identification of tissue-specific protein complexes in *Caenorhabditis elegans*.

[1] Co-first authorship.

SAMENVATTING VOOR NIET-INGEWIJDEN

Eiwit-eiwitinteracties & genoommodificatie

Nieuwe strategieën voor het bestuderen van celpolariteit

Een gepolariseerde cel is een cel waarvan de inhoud asymmetrisch is verdeeld en/of de vorm asymmetrisch is. Celpolariteit is een fundamentele eigenschap van cellen. Tal van cellulaire processen zijn erop gebaseerd, van asymmetrische celdeling tot het functioneren van epitheliale weefsels en zenuwen. Het verlies van celpolariteit is kenmerkend voor verschillende ziektes waaronder kanker.

Regulatoire eiwitten induceren en stabiliseren de polariteit van een cel in verschillende weefsels in vele organismen. Een aantal geconserveerde regulatoren van celpolariteit zijn bekend, waaronder de leden van de PAR-, Crumbs- en Scribble-eiwitgroepen. De regulatie van celpolariteit door algemene polariteitsregulatoren leidt tot een aantal vragen. Hoe polariseert dezelfde groep eiwitten een breed scala aan celtypes? Welke eiwitten interacteren met de geconserveerde regulatoren om polariteit tot stand te brengen? Om daar inzicht in te krijgen hebben we een weefselspecifieke eiwitzuiveringsmethode ontwikkeld voor de worm *Caenorhabditis elegans* (*C. elegans*). Dit modelorganisme wordt veel gebruikt voor onderzoek, omdat het makkelijk in een laboratorium te hanteren is en de mechanismen waarmee bijvoorbeeld polariteit tot stand komt goed vergelijkbaar zijn met de situatie in de mens. In de strategie die wij hebben ontwikkeld wordt zuivering van een bepaald eiwit uit één type weefsel gevolgd door identificatie van de interacterende eiwitten met massaspectrometrie (**hoofdstuk 1**). Biotine ligase BirA kan een biotine molecuul koppelen aan een eiwit dat gelabeld is met een Avi sequentie. Als BirA alleen aanwezig is in een bepaald weefsel, kunnen gelabelde eiwitten alleen in dat weefsel gebiotinyleerd worden. Daardoor komt de weefselspecificiteit van onze strategie tot stand. Streptavidine bindt aan biotine waardoor gebiotinileerde eiwitten met interacterende eiwitten gezuiverd kunnen worden met streptavidine-beklede bolletjes. De gebonden eiwitten worden vervolgens geïdentificeerd met massaspectrometrie. We hebben de weefselspecificiteit van de biotinylatie en zuivering aangetoond en onze aanpak toegepast op een aantal polariteitseiwitten, waarbij we gekeken hebben naar de darm en een epidermaal weefsel. Bij de zuivering van de reeds bekende polariteitsregulatoren DLG-1 uit de darm en CDC-42 uit epidermale cellen zijn interacterende eiwitten gevonden waarvan de interactie al beschreven was in literatuur. Deze resultaten laten zien dat deze strategie kan resulteren in het vinden van eiwit-eiwitinteracties in specifieke weefsels.

Naast het vinden van eiwitten die betrokken zijn bij de regulatie van celpolariteit in verschillende weefsels, is het van belang om een gedetailleerd beeld te krijgen van de mechanismen waarmee deze eiwitten een cel polariseren. Om het functioneren van een eiwit beter te begrijpen is het belangrijk te weten welke domeinen van het eiwit verantwoordelijk zijn voor welke functies. De functies van een eiwit worden voor een groot deel bepaald door de interacties die het aangaat met andere eiwitten. We hebben een methode waarmee eiwitdomeinen die interacteren met andere eiwitten in kaart worden gebracht opgeschaald tot genoomschaal (**hoofdstuk 2**). We hebben een verzameling genfragmenten gemaakt

van humane genen met behulp van ultrasonicatie en de kwaliteit van deze verzameling aangetoond door interacties van de corresponderende eiwitfragmenten met een set van polariteits- en celdelingseiwitten te testen. We vonden 22 eiwitinteracties waarvan 14 nieuwe interacties en acht interacties die al eerder beschreven zijn. Twaalf (55%) van de gevonden eiwitinteracties hebben we gevalideerd met een alternatieve aanpak, te weten affiniteitspurificatie in celkweek. Van de twaalf gevalideerde interacties waren zes interacties eerder beschreven en zes interacties nieuw gevonden. Dit toont aan dat deze strategie resulteert in een verzameling genfragmenten van hoge kwaliteit, die geschikt is om eiwit-eiwitinteracties en interactiedomeinen mee te identificeren.

Na het in kaart brengen van het eiwitdomein dat een bepaalde functie lijkt uit te voeren, bijvoorbeeld door aan een ander eiwit te binden, is het belangrijk om deze hypothese te testen in de context van een levend organisme. De beste aanpak voor zulke testen is het modificeren van het genoom van het betreffende organisme, zodat een gemuteerde versie van een eiwit ontstaat. Deze gemuteerde versie kan bijvoorbeeld bepaalde eiwitdomeinen missen. Andere opties zijn een stuk van het genoom vervangen door een andere sequentie of een stuk DNA toevoegen aan het genoom. Deze laatste optie maakt het mogelijk om eiwitten te voorzien van een fluorescent label, waarmee de dynamiek van het eiwit kan worden bestudeerd. Tot vorig jaar was het niet mogelijk om exacte modificaties in het genoom van *C. elegans* aan te brengen. Om het *C. elegans* genoom specifiek te kunnen modificeren hebben we een CRISPR/Cas9 strategie ontwikkeld (**hoofdstuk 3**). Het eiwit Cas9 wordt door een RNA molecuul gestuurd naar een specifieke plek in het genoom, waar dit eiwit een breuk maakt in het DNA. Door onnauwkeurige reparatie van de breuk kunnen mutaties ontstaan. We hebben mutaties in alle vier de geteste genen gevonden en daarmee laten zien dat deze strategie voor genoommodificatie in *C. elegans* werkt. Andere onderzoeksgroepen hebben onafhankelijk een CRISPR/Cas9 strategie ontwikkeld voor *C. elegans*. We hebben de overeenkomsten en verschillen van de ontwikkelde strategieën beschreven in **hoofdstuk 4**.

We hebben een variant van onze CRISPR/Cas9 strategie toegepast om de functie van het celpolariteitseiwit Crumbs te bestuderen in *C. elegans* (**hoofdstuk 5**). Hoewel het Crumbs eiwit essentieel is voor celpolariteit in epithelia van de fruitvlieg en daarmee voor diens levensvatbaarheid, lijken de twee tot dusver bekende *C. elegans* homologen van Crumbs geen cruciale rol te spelen bij de regulatie van celpolariteit in epithelia. We hebben een derde Crumbs homoloog gevonden in *C. elegans*. Dit eiwit is aanwezig in een gepolariseerd patroon in verschillende weefsels, net als in de fruitvlieg. Met een variant van onze CRISPR/Cas9 aanpak hebben we een drievoudige mutant gemaakt door de gehele genen van de drie Crumbs homologen te verwijderen. *C. elegans* mutanten die alle drie de Crumbs homologen missen zijn levensvatbaar, wat aangeeft dat de Crumbs homologen in dit modelorganisme geen essentiële rol spelen bij de regulatie van celpolariteit in tegenstelling tot de situatie in de fruitvlieg. Een overmaat van de drie *C. elegans* Crumbs homologen veroorzaakt een verandering in het localisatiepatroon van het polariteitseiwit PAR-3. Deze observaties

suggereren dat de *C. elegans* Crumbs homologen een niet-essentiële rol spelen bij de regulatie van celpolariteit.

In dit proefschrift zijn drie nieuwe strategieën beschreven die we ontwikkeld hebben voor het bestuderen van celpolariteit. Deze strategieën hebben al geresulteerd in nieuwe inzichten in de regulatie van celpolariteit en zullen daar ook in de toekomst aan bijdragen. Met name genoommodificatie met het CRISPR/Cas9 systeem zal impact hebben op eiwitstudies in elk proces.

DANKWOORD

Er zijn veel mensen die hebben bijgedragen aan mijn promotietijd, zowel op het lab als daarbuiten. Bedankt allemaal! De onderstaande mensen wil ik daar persoonlijk voor bedanken.

Alle collega's van de vakgroep Ontwikkelingsbiologie wil ik graag bedanken voor hun input en betrokkenheid. Mike, je hebt je vol ingezet om mij te begeleiden en daarvoor ben ik je dankbaar. Ik kon altijd bij je aankloppen voor overleg, face to face of digitaal. Ik heb een hoop van je geleerd de afgelopen jaren, onder andere om door te gaan, ook als het tegenzit (na een kort baalmomentje dan). Ook bedankt voor je hulp bij het schrijven. Dit proefschrift is vele malen beter geworden door jouw opmerkingen en aanvullingen. En ik ben blij dat ik wat langer mag blijven om het een en ander af te ronden en over te dragen. Sander, bedankt dat je mijn promotor bent. Jouw adviezen hebben een waardevolle invloed gehad op mijn onderzoek en proefschrift. Zoals op elke lab was het tijdens mijn AiO-tijd een komen en gaan van collega's. Slechts enkelen zijn getuige geweest van mijn promotietijd van begin tot eind. Adri, vanaf het begin leidde een kopje thee bij jou op kantoor tot nieuwe inzichten doordat je de situatie van een andere kant belichtte. Bedankt voor je bereidheid om met me mee te denken en voor het nuttige commentaar bij mijn manuscripten. Inge, het was leuk samen met je op het lab te staan. We werden ieder jaar heel gastvrij ontvangen door jou en Sander bij de thanksgiving feesten en barbeques. Vincent, bedankt dat je mijn vraagbaak was door de jaren heen en voor je hulp bij het maken van de cover van dit proefschrift. Je pogingen om mijn muziekkennis op te vijzelen hebben minder mooie resultaten opgeleverd. Suzan, het was fijn om mijn AiO-tijd vrijwel gelijktijdig met jou te doorlopen. We hebben vele experimenten door- en overdacht. Je was goed gezelschap op het lab en op congres. Thijs, je was een welkome versterking van Boxem & co. Jouw aanwezigheid zorgt voor leven in de brouwerij. Onze soepele samenwerking voor het Y2H hoofdstuk was erg plezierig. Jana, je hebt vele constructen voor mijn onderzoek gekloneerd, dank daarvoor. Daarnaast kun je ook lekkere baksels fabriceren, vooral de Ovocny kolac. Suzanne, tijdens onze skate-expedities (gelukkig vonden we altijd, uiteindelijk, de weg terug) en choco-momentjes hebben we heel wat afgebabbeld. Je aanstekelijke lach klonk geregeld over het lab en fleurde de boel op. Lars, je bent een gezellige noot op het lab en zit vol met mooie verhalen. Herman, het was leuk om een mede vogel-spotter op het lab te hebben en ik moet zeker m'n meerdere in jou erkennen. Martin, hoewel je geen echte OB-er bent, voelde dat wel zo. Bedankt voor je hulp en ik bewonder je wetenschappelijke enthousiasme. Ruben, jouw (zelf)spot heb ik gewaardeerd. João, it is good to have you back. Suzan, Thijs, Suzanne, Vincent, Lars, Ruben, and João, good luck with starting up/proceeding/finishing your PhDs. It was fun to waterski, kano by night, powerkite, watch soccer matches, have dinner, play games and go out with you (and others). Ook de oude garde collega's wil ik bedanken. Matilde en Christian, bedankt dat jullie mij wegwijs hebben gemaakt in de wereld van de wormen. Matilde, jouw optimisme en wetenschappelijke nieuwsgierigheid waren inspirerend. Ik vind het leuk dat we nog contact hebben en dat we welkom waren bij jullie in SF. Christian, je hebt een prachtig gevoel voor humor. Je grapjes uit onverwachte hoek hebben menigmaal voor vertier gezorgd onder het genot van een biertje (of meer).

En het was nog gezelliger met Emine erbij. Martine, Tim, Marjolein, Monique en Jerome, bedankt voor de leuke tijd. Alle stage-studenten, bedankt voor jullie bijdrage. Elisabeth, thank you for all your hard work. You were a great student and it was an honour to deliver your laudatio. I wish you a bright future. Alle co-auteurs, bedankt voor de samenwerking. Javier, Soenita, and Albert, thank you for the collaboration on the mass spec project, for all the high-quality mass spec runs and data analysis. Frits en Anja, bedankt dat jullie me op sleeptouw hebben genomen tijdens jullie wekelijkse hardlooprondjes. Mijn deelname aan de gesprekken nam af naarmate het rondje vorderde, gelukkig kletsten jullie vrolijk verder. Jammer dat een knieblessure en steeds drukkere werkdagen roet in het eten gooide toen ik jullie eindelijk een heel rondje bij kon houden. Everyone from Cell Biology, thanks for the nice retreats, Christmas parties, and chats in the hallway.

Naast (oud)collega's, wil ik ook graag een aantal vrienden en familieleden bedanken. Sanne, Sanne, Gisela en Els, onze klaverjasavonden staan garant voor een flinke portie lachen, maar ook serieuze (en) AiO perikelen worden gedeeld. Hopelijk zetten we de traditie nog lang voort! Sanne, jouw oprechte vriendschap in mooie en mindere tijden is me erg dierbaar. We kunnen alles bespreken en ik denk nog geregeld terug aan onze fantastische reis door Nieuw Zeeland. Dankjewel dat je mijn paranimf wilt zijn. Sanne, onze avondjes stamkroeg met Weihenstephaner mis ik nu al. Heel veel succes en plezier met je postdoc in Boston. En wel terugkomen, hè? Gisela, ik vind onze thee/lunch/eet-afspraakjes iedere keer weer zo gezellig. Binnenkort weer eens doen? Jolien, Ellen, Sophie, Janneke en Annejet, aka presto-meiden, we zijn als roeiploeg begonnen, maar dat bleek voor de meesten niet onze grootste hobby, koken en lekker eten daarentegen wel. Ik ben blij dat we nog steeds vriendinnen zijn en dat we de groeiende afstanden overbruggen om elkaar te zien en voor onze sinterklaasvieringen, kerstdiners en jaarlijkse weekendjes weg. Marlyne, wat zijn we al lang goede vriendinnen. Ik vind het erg leuk om samen met jou cultuur te snuiven tijdens Springdance of de Parade of om gewoon samen te eten. Mandy, je nuchtere kijk op de dingen is vaak verfrissend. Laten we als je terug bent uit Canada snel weer afspreken. Hsin-Yi, thank you for your friendship and I wish you all the best for the future. Rens en Annet, leuk dat we elkaar geregeld zien, spontaan of afgesproken. Jammer dat we binnenkort geen huis- en tuinbuurtjes meer zijn. An en Piet, bedankt voor jullie belangstelling in het verloop van mijn promotietraject. We mogen bij jullie steevast genieten van de kookkunsten. Inmiddels maak ik graag meerdere gerechten van Ans hand. Jeroen, Thera, Pieter-Eric, Lukas en Levi, het is leuk om geregeld met jullie bij te kletsen over werkgerelateerde en persoonlijke dingen. Riet en Joop, we hebben een bijzondere band. Ik vind het altijd gezellig met jullie. Er is dan ruimte voor zowel een geintje als voor een goed gesprek. Oma, tijdens mijn eerste stage heb je door de microscoop naar fluorescerende zebravisjes gekeken en stelde daar een paar pientere vragen over. Ik vind het heel speciaal dat je ook bij mijn promotiedag zal zijn. Max, grote broer, Kasia en Lieke, bedankt voor alle gezelligheid en uitjes de afgelopen jaren. Het voelde dan ongedwongen en ontspannen, of we nou gingen kanoën, wandelen, fietsen of naar een jazzcafé. En we worden vaak getrakteerd op één van Kasia's culinaire hoogstandjes. Max, bedankt dat je vandaag aan mijn zijde staat. Pap en mam, naast jullie interesse in mijn wel en wee, luisterend oor en adviezen (serieus of met een knipoog), ben ik jullie erg

dankbaar voor het hechte, warme nest dat jullie mij gegeven hebben. Jullie geven me het gevoel dat jullie trots op me zijn en staan altijd voor me klaar. Zo waren jullie zeer toegewijde niet-ingewijden bij het lezen van de Nederlandse samenvatting van dit proefschrift. Bas, dankjewel voor alle ruimte en support die je me de afgelopen jaren hebt gegeven. De successen van mijn AiO tijd hebben we samen gevierd, maar ook als het even tegen zat was je er altijd voor me. Onze tripjes naar het lab in de avond- of weekenduren, hebben ervoor gezorgd dat jij nu ook fluorescerende wormpjes kunt maken. Ik heb ontzettend veel zin in onze volgende grote reis binnenkort en al het andere wat we samen gaan beleven.

Selma