

Minimal Arrestin Domain Containing Protein 1 (ARRDC1) Constructs

Background of the Invention

Arrestin domain containing protein 1 (ARRDC1) drives the formation of extracellular
5 vesicles known as ARRDC1-mediated microvesicles (ARMMs), and these vesicles can be
harnessed to package and deliver a variety of molecular cargos, such as small molecules,
proteins, and nucleic acids.

Summary of the Invention

10 The arrestin domain containing protein 1 (ARRDC1) drives the formation of
extracellular vesicles, known as ARMMs (Nabhan J et al., PNAS 2012) and these vesicles
can be harnessed to package and deliver a variety of cargos, such as proteins, nucleic acids,
and small molecules (Wang Q and Lu Q, Nat Commun 2018). Full-length ARRDC1 protein
(433 amino acids at ~46 kD) has been used to recruit the molecular cargos into the vesicles,
15 either through a direct fusion with the molecular cargo or via a protein-protein interaction.
As ARRDC1 protein itself is packaged into ARMMs, and because the size of the vesicles is
limited (~80-100 nm), a smaller ARRDC1 protein that can still function in driving budding
would potentially increase the number of cargos that can be packaged into the vesicles.
Moreover, a smaller ARRDC1 may allow the recruitment of relatively large molecular
20 cargos.

Disclosed herein are minimal ARRDC1 proteins sufficient to drive ARMM budding.
The ARMM delivery system, described herein, addresses many limitations of current
delivery systems that prevent the safe and efficient delivery of molecular cargos, such as
small molecules, proteins, and nucleic acids to cells. As ARMMs are derived from an
25 endogenous budding pathway, they are unlikely to elicit a strong immune response, unlike
viral delivery systems, which are known to trigger inflammatory responses (Sen D. *et al.*,
“Cellular unfolded protein response against viruses used in gene therapy”, *Front*
Microbiology. 2014; 5:250, 1-16.). Additionally, ARMMs allow for the specific packaging
of any cargo protein of interest (*e.g.*, a targeted endonuclease such as a Cas9 protein, or Cas9
30 variant, with a guide RNA (gRNA)). These cargos can then be delivered by fusion or uptake
by specific recipient cells/tissues by incorporating antibodies or other types of molecules in
ARMMs that recognize tissue-specific markers. In some aspects, targeted endonucleases
such as Cas9-gRNAs and their variants can be loaded into ARMMs for delivery to a target cell.

ARMs are microvesicles that are distinct from exosomes and, like budding viruses, are produced by direct plasma membrane budding (DPMB). DPMB is driven by a specific interaction of TSG101 with the tetrapeptide PSAP (SEQ ID NO: 122) or PTAP (SEQ ID NO: 123) motif of the arrestin-domain-containing protein ARRDC1 accessory protein, which is

5 localized to the plasma membrane through its arrestin domain. ARMMs have been described in detail, for example, in PCT application number PCT/US2013/024839, filed February 6, 2013 (published as WO 2013/119602 A1) by Lu Q. *et al.*, and entitled “Arrdc1-mediated microvesicles (ARMMs) and uses thereof,” the entire contents of which are incorporated herein by reference; U.S. application number 14/929177, filed October 30, 2015 (published

10 as US 20160206566 A1) by Lu Q. *et al.*, entitled “Delivery of Cas9 via ARRDC1-Mediated Microvesicles (ARMMs),” the entire contents of which are incorporated herein by reference; and in PCT application number PCT/US2017/054912, filed October 03, 2017 (published as WO 2018/067546 A1) by Lu Q. *et al.*, and entitled “Delivery of Therapeutic RNAs via ARRDC1-Medicated Microvesicles,” the entire contents of which are incorporated herein by

15 reference. The ARRDC1/TSG101 interaction results in relocation of TSG101 from endosomes to the plasma membrane and mediates the release of microvesicles that contain TSG101, ARRDC1, and other cellular components.

Accordingly, in some embodiments, the present disclosure provides a minimal arrestin domain-containing protein 1 (ARRDC1) comprising an arrestin domain, at least one

20 PSAP (SEQ ID NO: 122) or PTAP (SEQ ID NO: 123) motif, and at least one PPXY motif, wherein the minimal ARRDC1 is shorter than full-length ARRDC1 protein. In some embodiments, the minimal ARRDC1 comprises at least two PPXY motifs. In some embodiments, the minimal ARRDC1 is less than 400 amino acids in length. In some embodiments, one or more of the PPXY motifs is PPEY (SEQ ID NO: 124). In some

25 embodiments, one or more of the PPXY motifs is PPSY (SEQ ID NO: 115). In some embodiments, at least two PPXY motifs are PPEY (SEQ ID NO: 124) and PPSY (SEQ ID NO: 115). In some embodiments, the minimal ARRDC1 comprises the amino acid sequence set forth in SEQ ID NO: 1.

Aspects of the present disclosure provide arrestin domain-containing protein 1

30 (ARRDC1)-mediated microvesicles (ARMM) comprising a lipid bilayer and a minimal ARRDC1 protein or variant thereof. In some embodiments, the minimal ARRDC1 protein comprises at least one PSAP (SEQ ID NO: 122) or PTAP (SEQ ID NO: 123) motif and at least one PPXY motif, and wherein the minimal ARRDC1 protein is shorter than full-length ARRDC1 protein. In some embodiments, the minimal ARRDC1 protein comprises the

amino acid sequence set forth in SEQ ID NO: 1. In some embodiments, the microvesicle further comprises an cargo, for example, a nucleic acid, a protein, and/or a small molecule.

In some embodiments, the microvesicle further comprises a TSG101 protein or fragment thereof. In some embodiments, the TSG101 protein fragment comprises a TSG101
 5 UEV domain. In some embodiments, the cargo to be delivered is conjugated to the minimal ARRDC1 protein, the minimal ARRDC1 fragment, the TSG101 protein, or the TSG101 fragment. In some embodiments, the microvesicle further comprises an integrin, a receptor tyrosine kinase, a G-protein coupled receptor, or a membrane-bound immunoglobulin.

In some embodiments, the microvesicle comprises an agent selected from the group
 10 consisting of Cas9 protein or Cas9 protein variant, Oct4, Sox2, c-Myc, KLF4 reprogramming factor, p53, Rb (retinoblastoma protein), BRCA1, BRCA2, PTEN, APC, CD95, ST7, ST14, a BCL-2 family protein, a caspase, BRMS1, CRSP3, DRG1, KAI1, KISS1, NM23, a TIMP-family protein, a BMP-family growth factor, EGF, EPO, FGF, G-CSF, GM-CSF, a GDF-family growth factor, HGF, HDGF, IGF, PDGF, TPO, TGF- α , TGF- β , VEGF; a zinc finger
 15 nuclease, Cre recombinase, Dre recombinase, FLP recombinase, , Hin, Gin, Tn3, β -six, CinH, ParA, $\gamma\delta$, Bxb1, ϕ C31, TP901, TG1, ϕ BT1, R4, ϕ RV1, ϕ FC1, MR11, A118, U153, gp29, Cre, FLP, R, Lambda, HK101, HK022, pSAM2, CAS9 nuclease or other Cas9-like targeted endonucleases such as Cpf1, CasX, CasY, or Geo), Sp1, NF1, CCAAT, GATA, HNF, PIT-1, MyoD, Myf5, Hox, Winged Helix, SREBP, p53, CREB, AP-1, Mef2, STAT, R-SMAD, NF-
 20 κ B, Notch, TUBBY, NFAT, α 1 β 1 integrin, α 2 β 1 integrin, α 4 β 1 integrin, α 5 β 1 integrin, α 6 β 1 integrin, α L β 2 integrin, α M β 2 integrin, α IIb β 3 integrin, α V β 3 integrin, α V β 5 integrin, α V β 6 integrin, α 6 β 4 integrin, EGF receptor (ErbB family), insulin receptor, PDGF receptor, FGF receptor, VEGF receptor, HGF receptor, Trk receptor, Eph receptor, AXL receptor, LTK receptor, TIE receptor, ROR receptor, DDR receptor, RET receptor, KLG receptor, RYK
 25 receptor, MuSK receptor, rhodopsin-like receptor, the secretin receptor, metabotropic glutamate/pheromone receptor, cyclic AMP receptor, frizzled/smoothed receptor, CXCR4, CCR5, beta-adrenergic receptor, CA19-9, c-met, PD-1, CTLA-4, ALK, AFP, EGFR, Estrogen receptor (ER), Progesterone receptor (PR), HER2/neu, KIT, B-RAF, S100, MAGE, Thyroglobulin, MUC-1, and PSMA.

30 In some embodiments, the agent (payload or cargo) to be delivered is a nucleic acid. In some embodiments, the nucleic acid comprises an RNA. In certain embodiments, the RNA is an RNAi agent. The RNA could be a coding RNA, a non-coding RNA, an antisense RNA, an mRNA, a guide RNA, a small RNA, an siRNA, an shRNA, a microRNA, an snRNA, a snoRNA, a lincRNA, or a structural RNA, or an rRNA or ribozyme.

In some embodiments, the nucleic acid comprises a DNA. In some embodiments, the DNA comprises a retrotransposon sequence, a LINE sequence, a SINE sequence, a composite SINE sequence, or an LTR-retrotransposon sequence.

In some embodiments, the nucleic acid agent encodes a protein. In some
5 embodiments, the agent comprises a detectable label.

In some embodiments, the agent comprises a therapeutic agent.

In some embodiments, the agent is selected from the group consisting of enzymes, antibodies, a Fab, a Fab', a F(ab')₂, a Fd, a scFv, a Fv, a dsFv, diabodies, and affibodies. In some embodiments, the agent comprises a cytotoxic agent.

10 In some embodiments, the agent comprises a protein. In some embodiments, the agent comprises a transcription factor, a transcriptional repressor, a fluorescent protein, a kinase, a phosphatase, a protease, a ligase, or a recombinase.

In some embodiments the agent comprises a lipid, or a lipoprotein, or a glycoprotein, or a polysaccharide, or a lipopolysaccharide.

15 In some embodiments, the agent is covalently bound to the ARRDC1 protein or fragment thereof, or the TSG101 protein or fragment thereof. In some embodiments, the agent is conjugated to the ARRDC1 protein or fragment thereof or the TSG101 protein or fragment thereof via a linker. In some embodiments, the linker is a cleavable linker. In some
20 embodiments, the linker comprises a protease recognition site or a UV-cleavable moiety, a photocleavable linker, or other linker cleavable by a biological mechanism, chemical degradation of a covalent bond, dissociation of a non-covalent association, a thermally labile link, or pH labile link.

In some embodiments, the agent to be delivered is fused to at least one WW domain or variant thereof. In some embodiments, the agent comprises two, three, four or five WW
25 domains or variants thereof. In some embodiments, the WW domain is derived from a WW domain of the ubiquitin ligase WWP1, WWP2, Nedd4-1, Nedd4-2, Smurf1, Smurf2, ITCH, NEDL1, or NEDL2. In some embodiments, the WW domain comprises a sequence selected from the group consisting of SEQ ID NO: 6-14.

In some embodiments, the agent is a protein. In some embodiments, the agent
30 comprises a Cas9 protein. In some embodiments, the Cas9 protein or variant thereof comprises at least one nuclear localization sequence (NLS). In some embodiments, the microvesicle further comprises a guide RNA (gRNA). In some embodiments, the WW domain is fused to the N-terminus of the protein. In some embodiments, the WW domain is fused to the C-terminus of the protein. In some embodiments, the microvesicle does not

include an exosomal biomarker. In some embodiments, the microvesicle is negative for an exosomal biomarker.

In some embodiments, the exosomal biomarker is chosen from the group consisting of CD63, Lamp-1, Lamp-2, CD9, HSPA8, GAPDH, CD81, SDCBP, PDCD6IP, ENO1,
5 ANXA2, ACTB, YWHAZ, HSP90AA129, ANXA5, EEF1A1, YWHAE, PPIA, MSN, CFL1, ALDOA, PGK1, EEF2, ANXA1, PKM2, HLA-DRA, and YWHAB. In some embodiments, the microvesicle does not include or is negative for Lamp-1. In some embodiments, the microvesicle diameter is from about 30 nm to about 500 nm.

Aspects of the present disclosure disclose an arrestin domain-containing protein 1
10 (ARRDC1)-mediated microvesicle (ARMM) as described above, further comprising a Cas9 cargo protein. The Cas9 cargo protein may be linked to the minimal ARRDC1 protein. The ARRDC1 protein can be covalently linked to the Cas9 cargo protein. In some embodiments, the minimal ARRDC1 protein is linked to the Cas9 protein via a cleavable linker. The linker
15 may be a UV-cleavable linker, and could include a protease recognition site or other linker cleavable by a biological mechanism, chemical degradation of a covalent bond, dissociation of a non-covalent association, a thermally labile link, or pH labile link.

In some aspects, the present disclosure provides an arrestin domain-containing protein 1 (ARRDC1)-mediated microvesicle (ARMM) comprising a minimal ARRDC1 protein or
20 variant thereof, and a cargo protein, wherein the cargo protein is linked to the TSG101 protein or variant thereof. In some aspects the cargo protein is linked to the TSG101 protein or variant by expression as a cargo-TSG101 fusion protein.

In some aspects, the present disclosure provides an arrestin domain-containing protein 1 (ARRDC1)-mediated microvesicle (ARMM) comprising a minimal ARRDC1 protein or
25 variant thereof, and a targeted endonuclease, wherein the targeted endonuclease is linked to the TSG101 protein or variant thereof.

In some aspects, the present disclosure provides an arrestin domain-containing protein 1 (ARRDC1)-mediated microvesicle (ARMM) comprising a minimal ARRDC1 protein or
variant thereof, and a Cas9 cargo protein, wherein the Cas9 cargo protein is linked to the TSG101 protein or variant thereof.

30 Some aspects of the present disclosure relate to minimal ARRDC1 fusion proteins. In some embodiments, the minimal ARRDC1 fusion protein comprises a minimal ARRDC1 protein or a variant thereof, wherein the minimal ARRDC1 protein comprises an arrestin domain, at least one PSAP (SEQ ID NO: 122) motif, at least one PPXY motif, and a Cas9 protein or variant thereof.

In some aspects, the present disclosure provides microvesicle-producing cells comprising a recombinant expression construct encoding a minimal ARRDC1 protein under the control of a heterologous promoter. In some embodiments, the microvesicle-producing cells further comprise a recombinant expression construct encoding a cargo protein under the control of a heterologous promoter. The cargo protein can be fused to at least one WW domain or variant thereof. The cargo protein can be expressed as a fusion protein with the minimal ARRDC1. The cargo protein can be expressed as a fusion protein with TSG101. In some embodiments, the microvesicle-producing cells further comprise a recombinant expression construct encoding a cargo nucleic acid produced under the control of a heterologous promoter.

In some aspects, the present disclosure includes a microvesicle-producing cell comprising a recombinant expression construct encoding a minimal ARRDC1 protein under the control of a heterologous promoter, wherein the minimal ARRDC1 protein comprises an arrestin domain, at least one PSAP (SEQ ID NO: 122) motif or PTAP (SEQ ID NO: 123) motif, and at least one PPXY motif, and wherein the minimal ARRDC1 protein is shorter than full-length ARRDC1 protein; and wherein the minimal ARRDC1 protein is linked to a Cas9 cargo protein or variant thereof. In some embodiments, the microvesicle producing cell comprises minimal ARRDC1 including at least two PPXY motifs.

Aspects of the present disclosure include methods of delivering a cargo to a target cell by contacting the target cell with any of the microvesicles or microvesicle-producing cells disclosed herein. Some aspects of the present disclosure provide methods of gene editing comprising contacting the target cell with any of the microvesicles or microvesicle-producing cells disclosed herein.

Other advantages, features, and uses of the invention will be apparent from the detailed description of certain exemplary, non-limiting embodiments; the drawings; the non-limiting working examples; and the claims.

Brief Description of the Drawings

FIG. 1 is a schematic showing the domain/motifs in full-length ARRDC1 protein and an example of a minimal ARRDC1.

FIGS. 2A-2C – **FIG. 2A** is a schematic showing the domain/motifs in full-length ARRDC1, an example of a minimal ARRDC1, and an ARRDC1 with a short N-terminus. **FIG. 2B** shows an image of a Western Blot. The constructs (all fused to the GFP protein) were transfected into HEK293T cells. 48 hours post transfection, extracellular vesicles (EVs)

were isolated via ultracentrifugation and used for anti-GFP Western blotting. Whole cell lysates of transfected cells were included. **FIG. 2C** is a graph showing the number of EVs in cultured media from transfected cells (as in **FIG. 2B**) as measured by a Nanosight NS300 instrument.

- 5 **FIGS. 3A-3C** – **FIG. 3A** is a schematic showing fusion of a minimal ARRDC1 to Cas9. **FIG. 3B** shows the results of anti-Flag Western blotting. Cas9, ARRDC1-Cas9, and miniARRDC1-Cas9 (all with a Flag tag at the C-terminus) were transfected into HEK293T cells. 48 hours post transfection, EVs were isolated via ultracentrifugation and used for anti-Flag Western blotting. Whole cell lysates of transfected cells were included. **FIG. 3C** is a
- 10 graph showing the gRNA ratio (ARMMs/cell) for Cas9/GFP-gRNA, a minimal ARRDC1-Cas9/GFP-gRNA, and a control. Cas9 and miniARRDC1-Cas9 along with a guide RNA (gRNA) that targets the GFP gene were transfected into donor HEK293T cells. 48 hours post transfection, EVs were isolated via ultracentrifugation. gRNA amount in ARMMs was measured by qRT-PCR and normalized against the cellular gRNA.
- 15 **FIG. 4** shows gene editing activity of miniARRDC1-Cas9. Cas9, ARRDC1-Cas9, and miniARRDC1-Cas9 along with a guide RNA (gRNA) targeting a single copy GFP gene transfected into U2OS cells. 48 hours post transfection, genomic DNAs were isolated and the GFP fragment containing the gRNA targeting site was amplified and analyzed by the T7E1 cleavage assay to detect indels (insertion or deletions).
- 20 **FIG. 5** shows the protein sequence of full-length ARRDC1 and an example of a minimal ARRDC1.

Definitions

- ARRDC1*: ARRDC1 is a protein that comprises a PSAP (SEQ ID NO: 122) and a
- 25 PPXY motif, also referred to herein as a PSAP (SEQ ID NO: 122) and PPXY motif, respectively, in its C-terminus, and interacts with TSG101 as shown herein. Exemplary, non-limiting ARRDC1 protein sequences are provided herein, and additional, suitable ARRDC1 protein variants according to aspects of this invention are known in the art. In addition, exemplary, non-limiting minimal ARRDC1 protein sequences are provided herein. It will be
- 30 appreciated by those of skill in the art that this invention is not limited in this respect. Exemplary ARRDC1 sequences include the following (PSAP (SEQ ID NO: 122) and PPXY motifs are marked):

>gil22748653|reflNP_689498.1| arrestin domain-containing protein 1 [*Homo sapiens*]

MGRVQLFEISLSHGRVVYSPGEPLAGTVRVRLGAPLPFRAIRVTCIGSCGVS NKANDT
 AWWVEEGYFNSSLSLADKGSLPAGEHSFPFQFLLPATAPTSFEGPFGKIVHQVRAAIH
 TPRFSKDHKCSLVFYILSPLNLNSIPDIEQPNVASATKKFSYKLVKTGSSVLTASTDLR
 GYVVGQALQLHADVENQSGKDTSPVVASLLQKVS YKAKRWIHDVRTIAEVEGAGV
 5 KAWRRAQWHEQILVPALPQSALPGCSLIHIDYYLQVSLKAPEATVTLVPVFIGNIAVNH
 APVSPRPGGLPPGAPPLVV[PSAP]PQEEAEAEAAAGGPHFLDPVFLSTKSHSQRQPLL
 ATLSSVPGAPEPCPDGSPASHPLHPPLCISTGATVPYFAEGSGGPVPTTSTLIL[PPEY]S
 SWGYPYEAPPSYEQSCGGVEPSLTPE (SEQ ID NO: 15)

10 >gil244798004|reflNP_001155957.1| arrestin domain-containing protein 1 isoform a [*Mus musculus*]
 MGRVQLFEIRLSQGRVVYGPGEPLAGTVHLRLGAPLPFRAIRVTCMGSCGVSTKAND
 GAWVVEESYFNSSLSLADKGSLPAGEHNFPFQFLLPATAPTSFEGPFGKIVHQVRASI
 DTPRFSKDHKCSLVFYILSPLNLNSIPDIEQPNVASTTKKFSYKLVKTGNVVLASTDL
 15 RGYVVGQVLRLQADIENQSGKDTSPVVASLLQKVS YKAKRWIYDVRTIAEVEGTGV
 KAWRRAQWQEILVPALPQSALPGCSLIHIDYYLQVSMKAPEATVTLPLFVGNI AVN
 QTPLSPCPGRESSPGTSLVV[PSAP]PQEEAEAVASGPHFSDPVSLSTKSHSQQQPLSAP
 LGSVSVTTTEPWVQVGSPARHSLHPPLCISIGATVPYFAEGSAGPVPTTSALIL[PPEY]S
 SWGYPYEAPPSYEQSCGAAGTDLGLIPGS (SEQ ID NO: 16)

20 >gil244798112|reflNP_848495.2| arrestin domain-containing protein 1 isoform b [*Mus musculus*]
 MGRVQLFEIRLSQGRVVYGPGEPLAGTVHLRLGAPLPFRAIRVTCMGSCGVSTKAND
 GAWVVEESYFNSSLSLADKGSLPAGEHNFPFQFLLPATAPTSFEGPFGKIVHQVRASI
 25 DTPRFSKDHKCSLVFYILSPLNLNSIPDIEQPNVASTTKKFSYKLVKTGNVVLASTDL
 RGYVVGQVLRLQADIENQSGKDTSPVVASLLQVS YKAKRWIYDVRTIAEVEGTGVK
 AWRRAQWQEILVPALPQSALPGCSLIHIDYYLQVSMKAPEATVTLPLFVGNI AVNQ
 TPLSPCPGRESSPGTSLVV[PSAP]PQEEAEAVASGPHFSDPVSLSTKSHSQQQPLSAPL
 GSVSVTTTEPWVQVGSPARHSLHPPLCISIGATVPYFAEGSAGPVPTTSALIL[PPEY]SS
 30 WGYPYEAPPSYEQSCGAAGTDLGLIPGS (SEQ ID NO: 17)

The term “ARMM,” as used herein, refers to a microvesicle comprising an ARDC1 protein or variant thereof, and/or TSG101 protein or variant thereof. In some embodiments,

the ARRDC1 protein or variant thereof is a minimal ARRDC1 protein or variant thereof. In some embodiments, the ARMM is shed from a cell, and comprises a molecule, for example, a nucleic acid, protein, or small molecule, present in the cytoplasm or associated with the membrane of the cell. In some embodiments, the ARMM is shed from a transgenic cell comprising a recombinant expression construct that includes the transgene, and the ARMM comprises a gene product, for example, a transcript or a protein (*e.g.*, a cargo protein) encoded by the expression construct. In some embodiments, the protein encoded by the expression construct is a Cas9 fusion protein, or Cas9 cargo protein fused to at least one WW domain, or variant thereof, which may associate with the minimal ARRDC1 protein to facilitate loading of the Cas9 cargo protein into the ARMM. In some embodiments, the ARMM is produced synthetically, for example, by contacting a lipid bilayer within ARRDC1 protein, or variant thereof, in a cell-free system in the presence of TSG101, or a variant thereof. In other embodiments, the ARMM is synthetically produced by further contacting a lipid bilayer with HECT domain ligase, and VPS4a. In some embodiments, an ARMM lacks a late endosomal marker. Some ARMMs as provided herein do not include, or are negative for, one or more exosomal biomarker. Exosomal biomarkers are known to those of skill in the art and include, but are not limited to, CD63, Lamp-1, Lamp-2, CD9, HSPA8, GAPDH, CD81, SDCBP, PDCD6IP, ENO1, ANXA2, ACTB, YWHAZ, HSP90AA1, ANXA5, EEF1A1, YWHAE, PPIA, MSN, CFL1, ALDOA, PGK1, EEF2, ANXA1, PKM2, HLA-DRA, and YWHAB. For example, some ARMMs provided herein lack CD63, some ARMMs lack LAMP1, some ARMMs lack CD9, some ARMMs lack CD81, some ARMMs lack CD63 and Lamp-1, some ARMMs lack CD63, Lamp-1, and CD9, some ARMMs lack CD63, Lamp-1, CD81, and CD9, and so forth. Certain ARMMs provided herein may include an exosomal biomarker. Accordingly, some ARMMs may be negative for one or more exosomal biomarkers but positive for one or more other exosomal biomarkers. For example, such an ARMM may be negative for CD63 and Lamp-1, but may include PGK1 or GAPDH; or may be negative for CD63, Lamp-1, CD9, and CD81, but may be positive for HLA-DRA. In some embodiments, ARMMs include an exosomal biomarker, but at a lower level than those typically found in exosomes. For example, some ARMMs include one or more exosomal biomarkers at a level of less than about 1%, less than about 5%, less than about 10%, less than about 20%, less than about 30%, less than about 40%, or less than about 50% of the level of that biomarker found in exosomes. To give a non-limiting example, in some embodiments, an ARMM may be negative for CD63 and Lamp-1, include CD9 at a level of less than about 5% of the level of CD9 typically found in exosomes, and be positive for

ACTB. Exosomal biomarkers in addition to those listed above are known to those of skill in the art, and the invention is not limited in this regard.

Agent and agent to be delivered: As used herein, the term “agent” refers to a substance that can be incorporated in an ARMM, for example, into the liquid phase of the ARMM or into the lipid bilayer of the ARMM. The term “agent to be delivered” refers to any substance that can be delivered to a subject, organ, tissue, or cell. In some embodiments, the agent is an agent to be delivered to a target cell. In some embodiments, the agent to be delivered is a biologically active agent, *i.e.*, it has activity in a cell, biological system, and/or subject. For instance, a substance that, when administered to a subject, has a biological effect on that subject, is considered to be biologically active. In some embodiments, an agent to be delivered is a therapeutic agent. As used herein, the term “therapeutic agent” refers to any agent that, when administered to a subject, has a beneficial effect. The term “therapeutic agent” refers to any agent that, when administered to a subject, has a therapeutic, diagnostic, and/or prophylactic effect and/or elicits a desired biological and/or pharmacological effect. As used herein, the term “therapeutic agent” may be a nucleic acid that is delivered to a cell via its association with or inclusion into an ARMM. In certain embodiments, the agent to be delivered is a nucleic acid. In certain embodiments, the agent to be delivered is DNA. In certain embodiments, the agent to be delivered is RNA. In certain embodiments, the agent to be delivered is a peptide or protein. In some embodiments, the functional protein or peptide to be delivered into a cell is a transcription factor, a tumor suppressor, a developmental regulator, a reprogramming factor, a growth factor, a metastasis suppressor, a pro-apoptotic protein, a zinc-finger nuclease, transcription activator-like effector nuclease, Cas9 protein, or a recombinase. In some embodiments, the protein to be delivered is p53, Rb (retinoblastoma protein), BRCA1, BRCA2, PTEN, APC, CD95, ST7, ST14, a BCL-2 family protein, a caspase; BRMS1, CRSP3, DRG1, KAI1, KISS1, NM23, a TIMP-family protein, a BMP-family growth factor, EGF, EPO, FGF, G-CSF, GM-CSF, a GDF-family growth factor, HGF, HDGF, IGF, PDGF, TPO, TGF- α , TGF- β , VEGF; a zinc finger nuclease, Cre, Dre, or FLP recombinase. In certain embodiments, the agent to be delivered is a small molecule. In some embodiments, the small molecule is an FDA-approved drug. In some embodiments, the agent to be delivered is a diagnostic agent. In some embodiments, the agent to be delivered is useful as an imaging agent. In some of these embodiments, the diagnostic or imaging agent is, and in others it is not, biologically active.

Animal: As used herein, the term “animal” refers to any member of the animal kingdom. In some embodiments, “animal” refers to humans of either sex at any stage of

development. In some embodiments, “animal” refers to non-human animals at any stage of development. In certain embodiments, the non-human animal is a mammal (*e.g.*, a rodent, a mouse, a rat, a rabbit, a monkey, a dog, a cat, a sheep, cattle, a primate, or a pig). In some embodiments, animals include, but are not limited to, mammals, birds, reptiles, amphibians, fish, and worms. In some embodiments, the animal is a transgenic animal, genetically-engineered animal, or a clone. In some embodiments, the animal is a transgenic non-human animal, genetically-engineered non-human animal, or a non-human clone.

Approximately: As used herein, the term “approximately” or “about,” as applied to one or more values of interest, refers to a value that is similar to a stated reference value. In certain embodiments, the term “approximately” or “about” refers to a range of values that fall within 25%, 20%, 19%, 18%, 17%, 16%, 15%, 14%, 13%, 12%, 11%, 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, 1%, or less in either direction (greater than or less than) of the stated reference value unless otherwise stated or otherwise evident from the context (except where such number would exceed 100% of a possible value).

Associated with: As used herein, the terms “associated with,” “conjugated,” “linked,” “attached,” and “tethered,” when used with respect to two or more entities, for example, moieties, molecules, and/or ARMMs, means that the entities are physically associated or connected with one another, either directly or via one or more additional moieties that serve as a linker, to form a structure that is sufficiently stable so that the entities remain physically associated under the conditions in which the structure is used, *e.g.*, physiological conditions. An ARMM is typically associated with an agent, for example, a nucleic acid, protein, or small molecule, by a mechanism that involves a covalent or non-covalent association. In certain embodiments, the agent to be delivered is covalently bound to a molecule that is part of the ARMM, for example, an ARRC1 protein or fragment thereof, a TSG101 protein or fragment thereof, or a lipid or protein that forms part of the lipid bilayer of the ARMM. In some embodiments, a peptide or protein is associated with an ARRC1 protein or fragment thereof, a TSG101 protein or fragment thereof, or a lipid bilayer-associated protein by a covalent bond (*e.g.*, an amide bond). In some embodiments, the association is via a linker, for example, a cleavable linker. In some embodiments, an entity is associated with an ARMM by inclusion in the ARMM, for example, by encapsulation of an entity (*e.g.*, an agent) within the ARMM. For example, in some embodiments, an agent present in the cytoplasm of an ARMM-producing cell is associated with an ARMM by encapsulation of agent-comprising cytoplasm in the ARMM upon ARMM budding. Similarly, a membrane protein, or other molecule associated with the cell membrane of an ARMM producing cell may be associated

with an ARMM produced by the cell by inclusion into the ARMM membrane upon budding. In certain embodiments, the agent to be delivered comprises a WW domain that effects binding to proline containing or proline rich domains in a molecule that is part of the ARMM.

Biologically active: As used herein, the phrase “biologically active” refers to a characteristic of any substance that has activity in a biological system and/or organism. For instance, a substance that, when administered to an organism, has a biological effect on that organism, is considered to be biologically active. In particular embodiments, where a nucleic acid is biologically active, a portion of that nucleic acid that shares at least one biological activity of the whole nucleic acid is typically referred to as a “biologically active” portion. As one example, a nuclease cargo protein may be considered biologically active if it increases or decreases the expression of a gene product when administered to a subject.

Biomarker: The term “biomarker”, as used herein, in the context of ARMM-based diagnostics, refers to a detectable molecule (e.g., a protein, a peptide, a proteoglycan, a glycoprotein, a lipoprotein, a carbohydrate, a lipid, a nucleic acid (e.g., DNA, such as cDNA or amplified DNA, or RNA, such as mRNA), an organic or inorganic chemical compound, a small molecule (e.g., second messenger, a metabolite), or a discriminating molecule or discriminating fragment of any of the foregoing), that is present in or derived from a biological sample containing ARMMs, or any ratio of such molecules, or any other characteristic that is objectively measured and evaluated as an indicator of a specific biological feature or process, for example, of cell or vesicle identity, of a normal or a pathogenic processes, or a pharmacologic response to a therapeutic intervention, or an indication thereof. See Atkinson, A. J., *et al.*, Biomarkers and Surrogate Endpoints: Preferred Definitions and Conceptual Framework, *Clinical Pharm. & Therapeutics*, 2001 March; 69(3): 89-95. In this context, the term “derived from” refers to a compound that, when detected, is indicative of a particular molecule being present in the biological sample. For example, detection of a particular cDNA can be indicative of the presence of a particular RNA transcript or protein in the biological sample. As another example, detection of or binding to a particular antibody can be indicative of the presence of a particular antigen (e.g., protein) in the biological sample. In some embodiments, a discriminating molecule or fragment is a molecule or fragment that, when detected, indicates presence or abundance of the compound of which its presence is indicative. A biomarker can, for example, be isolated from an ARMM, directly measured as part of an ARMM, or detected in or determined to be included in an ARMM. In some embodiments, the amount of ARMMs detected in a sample from a subject or in a cell population derived from a sample obtained from a subject, or the

rate of ARMM generation within a sample or cell population obtained from a subject serves as a biomarker. Methods for the detection of biomarkers are known to those of skill in the art and include nucleic acid detection methods, protein detection methods, carbohydrate detection methods, antigen detection methods, small molecule detection methods, and other suitable methods.

Biomarker profile: A “biomarker profile” comprises one or more biomarkers (e.g., an mRNA molecule, a cDNA molecule, a protein, and/or a carbohydrate, or an indication thereof). The biomarkers of the biomarker profile can be in the same or different classes, such as, for example, a nucleic acid, a carbohydrate, a metabolite, and a protein. A biomarker profile may comprise at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, or more biomarkers. In some embodiments, a biomarker profile comprises hundreds, or even thousands, of biomarkers. A biomarker profile can further comprise one or more controls or internal standards. In some embodiments, the biomarker profile comprises at least one biomarker that serves as an internal standard. In some embodiments, the presence or level of ARRDC1 or TSG101 in a sample or cell population is used as an internal standard. Biomarker profiles for several conditions, diseases, and pathologies, and also for normal states are known to those of skill in the art, and the invention is not limited to any particular biomarker profile. In some embodiments, the biomarker profile used in the context of ARMM-based diagnostic methods as described herein is a biomarker profile that has been described to be useful for exosome-based diagnostics. Exosomal biomarker profiles are known to those of skill in the art and biomarker profiles useful for the diagnosis of various disease, including different cancers, stroke, autism, and other diseases, have been described, for example, in U.S. Patent Application, U.S.S.N. 13/009,285, filed on January 19, 2011 (published as US 2011/0151460 A1) by Kaas *et al.*, and entitled Methods And Systems Of Using Exosomes For Determining Phenotypes, the entire contents of which are incorporated herein by reference.

Cas9 or Cas9 Protein: The term “Cas9” or “Cas9 protein” refers to an RNA-guided nuclease comprising a Cas9 protein, or a variant thereof (e.g., a protein comprising an active, inactive, or altered DNA cleavage domain of Cas9, and/or the gRNA binding domain of Cas9). A Cas9 nuclease is also referred to sometimes as a casn1 nuclease or a CRISPR (clustered regularly interspaced short palindromic repeat)-associated nuclease. CRISPR is an adaptive immune system that provides protection against mobile genetic elements (viruses, transposable elements and conjugative plasmids). CRISPR clusters contain spacers, sequences complementary to antecedent mobile elements, and target invading nucleic acids.

CRISPR clusters are transcribed and processed into CRISPR RNA (crRNA). In type II CRISPR systems correct processing of pre-crRNA requires a trans-encoded small RNA (tracrRNA), endogenous ribonuclease 3 (rnc) and a Cas9 protein. The tracrRNA serves as a guide for ribonuclease 3-aided processing of pre-crRNA. Subsequently,

5 Cas9/crRNA/tracrRNA endonucleolytically cleaves linear or circular dsDNA target complementary to the spacer. The target strand not complementary to crRNA is first cut endonucleolytically, then trimmed 3'-5' exonucleolytically. In nature, DNA-binding and cleavage typically requires protein and both RNAs. However, single guide RNAs ("sgRNA", or simply "gRNA") can be engineered so as to incorporate aspects of both the crRNA and

10 tracrRNA into a single RNA species. See, *e.g.*, Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J.A., Charpentier E. *Science* 337:816-821(2012), the entire contents of which is hereby incorporated by reference. Cas9 recognizes a short motif in the CRISPR repeat sequences (the PAM or protospacer adjacent motif) to help distinguish self versus non-self. Cas9 nuclease sequences and structures are well known to those of skill in the art (see, *e.g.*,

15 "Complete genome sequence of an M1 strain of *Streptococcus pyogenes*." Ferretti *et al.*, J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K., Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P., Qian Y., Jia H.G., Najar F.Z., Ren Q., Zhu H., Song L., White J., Yuan X., Clifton S.W., Roe B.A., McLaughlin R.E., *Proc. Natl. Acad. Sci. U.S.A.* 98:4658-4663(2001); "CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III." Deltcheva E., Chylinski K., Sharma C.M., Gonzales K., Chao Y., Pirzada Z.A., Eckert M.R., Vogel J., Charpentier E., *Nature* 471:602-607(2011); and "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity." Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J.A., Charpentier E. *Science* 337:816-821(2012), the entire contents of each of which are incorporated herein by reference). Cas9

20 orthologs have been described in various species, including, but not limited to, *S. pyogenes* and *S. thermophilus*. Additional suitable Cas9 nucleases and sequences will be apparent to those of skill in the art based on this disclosure, and such Cas9 nucleases and sequences include Cas9 sequences from the organisms and loci disclosed in Chylinski, Rhun, and Charpentier, "The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems" (2013) *RNA Biology* 10:5, 726-737; the entire contents of which are incorporated herein by

30 reference. In some embodiments, a Cas9 nuclease has an inactive (*e.g.*, an inactivated) DNA cleavage domain.

A nuclease-inactivated Cas9 protein may interchangeably be referred to as a "dCas9" protein (for nuclease-"dead" Cas9). Methods for generating a Cas9 protein (or a variant

thereof) having an inactive DNA cleavage domain are known (See, *e.g.*, Jinek *et al.*, *Science*. 337:816-821(2012); Qi *et al.*, “Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression” (2013) *Cell*. 28;152(5):1173-83, the entire contents of each of which are incorporated herein by reference). For example, the DNA

5 cleavage domain of Cas9 is known to include two subdomains, the HNH nuclease subdomain and the RuvC1 subdomain. The HNH subdomain cleaves the strand complementary to the gRNA, whereas the RuvC1 subdomain cleaves the non-complementary strand. Mutations within these subdomains can silence the nuclease activity of Cas9. For example, the mutations D10A and H841A completely inactivate the nuclease activity of *S. pyogenes* Cas9

10 (Jinek *et al.*, *Science*. 337:816-821(2012); Qi *et al.*, *Cell*. 28;152(5):1173-83 (2013). In some embodiments, proteins comprising variants of Cas9 are provided. For example, in some embodiments, a protein comprises one of two Cas9 domains: (1) the gRNA binding domain of Cas9; or (2) the DNA cleavage domain of Cas9. In some embodiments, proteins comprising Cas9 or variants thereof are referred to as “Cas9 variants.” A Cas9 variant shares

15 homology to Cas9, or a variant thereof. For example, a Cas9 variant is at least about 70% identical, at least about 80% identical, at least about 90% identical, at least about 95% identical, at least about 96% identical, at least about 97% identical, at least about 98% identical, at least about 99% identical, at least about 99.5% identical, or at least about 99.9% to wild type Cas9. In some embodiments, the Cas9 variant comprises a variant of Cas9 (*e.g.*,

20 a gRNA binding domain or a DNA-cleavage domain), such that the variant is at least about 70% identical, at least about 80% identical, at least about 90% identical, at least about 95% identical, at least about 96% identical, at least about 97% identical, at least about 98% identical, at least about 99% identical, at least about 99.5% identical, or at least about 99.9% to the corresponding variant of wild type Cas9. In some embodiments, wild type Cas9

25 corresponds to Cas9 from *Streptococcus pyogenes* (NCBI Reference Sequence: NC_017053.1, SEQ ID NO: 108(nucleotide); SEQ ID NO: 2 (amino acid)).

ATGGATAAGAAATACTCAATAGGCTTAGATATCGGCACAAATAGCGTCGGATGGGCGGTGATCACTGATGATTAT
 AAGGTTCCGTCTAAAAAGTTCAAGGTTCTGGGAAATACAGACCGCCACAGTATCAAAAAAATCTTATAGGGGCT
 CTTTTATTTGGCAGTGGAGAGACAGCGGAAGCGACTCGTCTCAAACGGACAGCTCGTAGAAGGTATACACGTCGG
 30 AAGAATCGTATTTGTTATCTACAGGAGATTTTTTCAAATGAGATGGCGAAAGTAGATGATAGTTTCTTTTCATCGA
 CTTGAAGAGTCTTTTTTGGTGGGAAGAAGACAAGAAGCATGAACGTCATCCTATTTTTGGAAATATAGTAGATGAA
 GTTGCTTATCATGAGAAATATCCAACCTATCTATCATCTGCGAAAAAATTTGGCAGATTCTACTGATAAAGCGGAT
 TTGCGCTTAATCTATTTGGCCTTAGCGCATATGATTAAGTTTCGTGGTCATTTTTTGATTGAGGGAGATTTAAAT
 CCTGATAATAGTGATGTGGACAACTATTTATCCAGTTGGTACAAATCTACAATCAATTATTTGAAGAAAACCCCT
 35 ATTAACGCAAGTAGAGTAGATGCTAAAGCGATTCTTTCTGCACGATTGAGTAAATCAAGACGATTAGAAAATCTC
 ATTGCTCAGCTCCCCGGTGAGAAGAGAAATGGCTTGTTGGGAATCTCATTGCTTTGTCATTGGGATTGACCCCT
 AATTTTAAATCAAATTTTGATTGGCAGAAAGATGCTAAATTACAGCTTTCAAAGATACCTACGATGATGATTTA
 GATAATTTATTGGCGCAAATTTGGAGATCAATATGCTGATTTGTTTTTGGCAGCTAAGAATTTATCAGATGCTATT
 TTACTTTTCAGATATCCTAAGAGTAAATAGTGAAATAACTAAGGCTCCCTATCAGCTTCAATGATTAAGCGCTAC
 40 GATGAACATCATCAAGACTTGACTCTTTTAAAGCTTTAGTTTCGACAACAACCTCCAGAAAAGTATAAAGAAATC

TTTTTGGATCAATCAAAAAACGGATATGCAGGTTATATTGATGGGGGAGCTAGCCAAGAAGAATTTTATAAATTT
ATCAAAACCAATTTTAGAAAAATGGATGGTACTGAGGAATTATTGGTGAAACTAAATCGTGAAGATTTGCTGCGC
AAGCAACGGACCTTTGACAAACGGCTCTATCCCCATCAAATTCACCTGGGTGAGCTGCATGCTATTTGAGAAGA
CAAGAAGACTTTTATCCATTTTAAAAAGACAATCGTGAGAAGATTGAAAAATCTTGACTTTTCGAATTCCTTAT
5 TATGTTGGTCCATTGGCGCGTGGCAATAGTCGTTTTGCATGGATGACTCGGAAGCTGAAGAAACAATTACCCCA
TGAATTTTGAAGAAGTTGTCGATAAAAGGTGCTTCAGCTCAATCATTTATTGAACGCATGACAACTTTGATAAAA
AATCTTCCAAATGAAAAAGTACTACCAAAACATAGTTTGCTTTATGAGTATTTTACGGTTTATAACGAATTGACA
AAGGTCAAATATGTTACTGAGGGAATGCGAAAACCAGCATTTCCTTCAGGTGAACAGAAGAAAGCCATTGTTGAT
TTACTCTTCAAACAAATCGAAAAGTAACCGTTAAGCAATTAAGAAGATTATTTCAAAAAATAGAAATGTTTT
10 GATAGTGTGAAATTTTACAGGAGTTGAAGATAGATTTAATGCTTCATTAGGCGCTACCATGATTTGCTAAAAAT
ATTAAAGATAAAGATTTTGGATAATGAAGAAATGAAGATATCTTAGAGGATATTGTTTTAACATTGACCTTA
TTTGAAGATAGGGGATGATTGAGAAAAGACTTAAACATATGCTCACCTCTTTGATGATAAGGTGATGAAACAG
CTTAAACGTCGCCGTTTACTGTTGGGGACGTTTGCTCGAAAATTGATTAATGGTATTAGGGATAAGCAATCT
GGCAAAACAATATTAGATTTTGAATCAGATGGTTTTGCCAATCGCAATTTTATGCAGCTGATCCATGATGAT
15 AGTTTGACATTTAAAGAAGATATTCAAAGACAGGTGTCTGGACAAGGCCATAGTTTACATGAACAGATTGCT
AACTTAGCTGGCAGTCTGCTATTAATAAAGGTATTTTACAGACTGTAAAAATTGTTGATGAACCTGGTCAAAGTA
ATGGGGCATAAGCCAGAAAAATATCGTTATTGAAATGGCAGCTGAAAAATCAGACAACCTCAAAGGGCCAGAAAAAT
TCGCGAGAGCGTATGAAACGAATCGAAGAAGGTATCAAAGAATTAGGAAGTCAGATTCTTAAAGAGCATCCCTGTT
GAAAAACTCAATTGCAAAATGAAAAGCTCTATCTCTATTATCTACAAAATGGAAGAGACATGTATGTGGACCAA
20 GAATTAGATATTAATCGTTTAAAGTATTATGATGTCGATCACATTGTTCCACAAAGTTTCATTAAAGACGATTCA
ATAGACAATAAGGTACTAACGCGTTCTGATAAAAATCGTGGTAAATCGGATAACGTTCCAAGTGAAGAAGTACTC
AAAAAGATGAAAAACTATTGGAGACAACCTTCAAACGCCAAGTTAATCACTCAACGTAAGTTTGATAATTTAACG
AAAGCTGAACGTGGAGGTTTGAGTGAACCTTGATAAAGCTGGTTTTATCAAACGCCAATTGGTTGAAACTCGCCAA
ATCACTAAGCATGTGGCACAATAATTTGGATAGTCGCATGAATACTAAATACGATGAAAAATGATAAATTTATTCGA
25 GAGGTTAAAGTGATTACCTTAAATCTAAATAGTTTCTGACTTCCGAAAAGATTTCCAATTTATAAAGTACGT
GAGATTAACAATTACCATCATGCCCATGATGCGTATCTAAATGCCGTGCTTGGAACTGCTTTGATTAAGAAATAT
CCAAAACCTTGAATCGGAGTTTGTCTATGGTGATTATAAAGTTTATGATGTTCTGTAAGTATTGCTAAGTCTGAG
CAAGAAATAGGCAAGCAACCGCAAAATATTTCTTTTACTCTAATATCATGAACCTCTTCAAACAGAAATTACA
CTTGCAAAATGGAGGATTTCGCAACCGCCCTCTAATCGAACTAATGGGGAACCTGGAGAAATTTGCTGGGATAAA
30 GGGCGAGATTTTGCCACAGTGGCGCAAAGTATTGTCTCATGCCCCAAGTCAATATTGTCAAGAAAACAGAAAGTACAG
ACAGGCGGATTCTCCAAGGAGTCAATTTTACCAAAAAGAAATTCGGACAAGCTTATTGCTCGTAAAAAAGACTGG
GATCCAAAAAATATGGTGGTTTTGATAGTCCAACGCTAGCTTATTCAGTCCTAGTGGTTGCTAAGGTGGAAAAA
GGGAAATCGAAGAAGTTAAATCCGTTAAAGAGTTACTAGGGATCACAATTATGGAAGAAGTTCCCTTTGAAAAA
AATCCGATTGACTTTTGAAGCTAAAGGATATAAGGAAGTTAAAAAAGACTTAATCATTAACCTACCTAAATAT
35 AGTCTTTTGGATTGATAAAGCGGTGTAACGAGTGTGGCTAGTGGCGGAGAATTACAAAAAGGAAATGAGCTG
GCTCTGCCAAGCAAAATATGTGAATTTTTTATATTTAGCTAGTCATTATGAAAAGTTGAAGGGTAGTCCAGAAGAT
AACGAACAAAAACAATTGTTTGTGGAGCAGCATAAGCATTATTTAGATGAGATTATTGAGCAAAATCAGTGAATTT
TCTAAGCGTGTATTTTAGCAGATGCCAATTTAGATAAAGTTCTTAGTGCATATAACAAACATAGAGACAAACCA
ATACGTGAACAAGCAGAAAAATATTATTCATTTATTTACGTTGACGAATCTTGGAGCTCCCGCTGCTTTTAAATAT
40 TTTGATACAACAATTGATCGTAAACGATATACGCTACAAAAGAAGTTTTAGATGCCACTCTTATCCATCAATCC
ATCACTGGTCTTTATGAAACACGCATTGATTTGAGTCAGCTAGGAGGTGACTGA (SEQ ID NO: 1)

MDKKYSIGLDIGTNSVGVAVITDDYKVPKSKFKVLGNTRHSIKKNLIGALLFGSGETAETRLKRTARRRYTRR
KNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLADSTDKAD
45 LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQIYNQLFEENPINASRVDAKAIL SARLSKSRRLENL
IAQLPGEKRNLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADFLAAKNLSDAI
LLSDILRVNSEITKAPLSAMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKF
IKPILEKMDGTEELLVKNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTRIPY
YVGPLARGNSRFAMWTRKSEETITPWNFEVVVDKGASQSFIERMTNFDKNLPNEKVLPHKSLLEYFTVYNELT
50 KVYVTEGMRKPAFLSGEQKAIVDLLFKTNRKVTQKLEDYFKKIECFDSVEISGVEDRFNASLGAYHDLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDRGMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQS
GKTILDFLKSDFANRNFMLIHDDSLTFKEDIQKAQVSGQGHSLHEQIANLAGSPAIIKKGIQTIVKIVDELVKV
MGHKPENIVIEEMARENQTTQKGQKNSRERMKRIEIEGIELGSQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQ
ELDINRLSDYDVIDHIVPQSFIKDDSIDNKVLTRSDKNRGSNDNPSEEVVKMKMNYWRQLLNKILITQRKFDNLT
55 KAERGGSELKAGFIKRLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVR
EINNYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVDVRKMIKSEQEI GKATAKYFFYSNIMNFFKTEIT
LANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIKKTEVQTGGFSKESILPKRNSDKLIARKKDW
DPKKYGGDFSPTVASVVLVAKVEKGSKKLKSVELLGHITIMERSSEFKNPIDFLEAKGYKEVKKDLIILPKPY
SLFELENGRKMALASAGELQKGNELALPSKYVNFYLASHYKELKGSPEDEQKQLFVEQHKHYLDEIIEQISEF
60 SKRVILADANLDKVL SAYNKHDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQ
ITGLYETRIDLSQLGGD (SEQ ID NO:2)

(single underline: HNH domain; double underline: RuvC domain)

In some embodiments, wild type (*S. pyogenes*) Cas9 corresponds to, or comprises

SEQ ID NO: 3 (nucleotide) and/or SEQ ID NO: 4 (amino acid):

```
5  ATGGATAAAAAGTATTCTATTGGTTTTAGACATCGGCACTAATTCGGTTGGATGGGCTGTCATAACCGATGAATAC
   AAAGTACCTTCAAAGAAATTTAAGGTGTTGGGGAACACAGACCGTCATTGATTAAAAAGAATCTTATCGGTGCC
   CTCTATTTCGATAGTGGCGAAACGGCAGAGGCGACTCGCCTGAAACGAACCGCTCGGAGAAGGTATACACGTGCG
   AAGAACCGAATATGTTACTTACAAGAAATTTTTAGCAATGAGATGGCCAAAGTTGACGATTCTTTCTTTACCGT
   TTGGAAGAGTCCTTCTTTGTCGAAGAGGACAAGAAACATGAACGGCACCCCATCTTTGGAAACATAGTAGATGAG
10  GTGGCATATCATGAAAAGTACCCAACGATTTATCACCTCAGAAAAAGCTAGTTGACTCAACTGATAAAGCGGAC
   CTGAGGTTAATCTACTTGGCTCTTGCCCATATGATAAAGTTCCGTGGGCACTTTCTCATTGAGGGTGATCTAAAT
   CCGGACAACCTCGGATGTCGACAACTGTTTCATCCAGTTAGTACAAACCTATAATCAGTTGTTTGAAGAGAACCCT
   ATAAATGCAAGTGGCGTGGATGCGAAGGCTATTCTTAGCGCCCGCCTCTCTAAATCCCGACGGCTAGAAAAACCTG
   ATCGCACAAATTACCCGGAGAGAAGAAAAATGGGTTGTTGCGTAACCTTATAGCGCTCTCACTAGGCGCTGACACCA
15  AATTTTAAAGTCGAACCTTCGACTTAGCTGAAGATGCCAAATTGCAGCTTAGTAAGGACACGTACGATGACGATCTC
   GACAATCTACTGGCACAAATTTGGAGATCAGTATGCGGACTTATTTTTGGCTGCCAAAAACCTTAGCGATGCAATC
   CTCTATCTGACATACTGAGAGTTAATACTGAGATTACCAAGGCGCGCTTATCCGCTTCAATGATCAAAAGGTAC
   GTGAACATCACCAAGACTTGACACTTCTCAAGGCCCTAGTCCGTGAGCAACTGCCTGAGAAATATAAGGAAATA
   TTCTTTGATCAGTCGAAAAACGGGTACGCAGGTTATATTGACGGCGGAGCGAGTCAAGAGGAATTTCTACAAGTTT
20  ATCAAAACCCATATTAGAGAAGATGGATGGGACGGAAGAGTTGCTTGTAAACTCAATCGCGAAGATCTACTGCGA
   AAGCAGCGGACTTTGACAAACGGTAGCATTCACATCAAATCCACTTAGGCGAATTGCATGCTATACTTAGAAGG
   CAGGAGGATTTTTATCCGTTCCTCAAAGACAATCGTGAAAAGATTGAGAAAATCCTAACCTTTTCGCATACCTTAC
   TATGTGGGACCCCTGGCCCGAGGGAACCTCGGTTTCGATGGATGACAAGAAAGTCCGAAGAAACGATTACTCCA
   TGGAATTTTGAGGAAGTTGTCGATAAAGGTGCGTCAGCTCAATCGTTTCATCGAGAGGATGACCAACTTTTGACAAG
25  AATTTACCGAACGAAAAAGTATTGCCTAAGCACAGTTTACTTTACGAGTATTTTACAGTGTACAAATGAACCTCAG
   AAAGTTAAGTATGTCAGTGGGCGATGCGTAAACCCGCTTTCTAAGCGGAGAACAGAAGAAAGCAATAGTAGAT
   CTGTTATTCAAGACCAACCGCAAAGTGACAGTTAAGCAATTGAAAGAGGACTACTTTAAGAAAATTGAATGCTTC
   GATTCTGTGCGATCTCCGGGTAGAAAGATCGATTTAATGCGTCACTTGGTACGTATCATGACCTCCTAAAGATA
   ATTAAGATAAGGACTTCTTGGATAACGAAGAGAATGAAGATATCTTAGAAGATATAGTGTGACTCTTACCCCTC
30  TTTGAAGATCGGGAATGATTGAGGAAAGACTAAAAACATACGCTCACCTGTTTCGACGATAAGGTTATGAAACAG
   TTAAAGAGGCGCTCGCTATACGGGCTGGGACGATTGTGCGGAAACTTATCAACGGGATAAGAGACAAGCAAGT
   GGTAAACTATTCTCGATTTTCTAAAGAGCGACGGCTTCGCCAATAGGAACCTTTATGCAGCTGATCCATGATGAC
   TCTTTAACCTTCAAAGAGGATATACAAAAGGCACAGGTTTCGGGACAAGGGGACTCATTGCACGAACATATTGCG
   AATCTTGCTGGTTTCGCCAGCCATCAAAAGGGCATACTCCAGACAGTCAAAGTAGTGGATGAGCTAGTTAAGGTC
35  ATGGGACGTCACAAACCGGAAAAACATTGTAATCGAGATGGCAGCGAAAAATCAAACGACTCAGAAGGGGCAAAAA
   AACAGTCGAGAGCGGATGAAGAGAATAGAAGAGGTTATTAAGAACTGGGCAGCGAGATCTTAAAGGAGCATCCT
   GTGGAAAAATCCCAATTGCAGACGAGAACTTTACCTCTATTACCTACAAAATGGAAGGACATGTATGTTGAT
   CAGGAACTGGACATAAACCGTTTATCTGATTACGACGTCGATCACATTGTACCCCAATCCTTTTTGAAGGACGAT
   TCAATCGACAATAAAGTGCTTACACGCTCGGATAAGAACCAGGGGAAAGTGACAAATGTTCCAAGCGAGGAAGTC
40  GTAAAGAAAATGAAGAACTATTGGCGGCGAGCTCCTAAATGCGAACTGATAACGCAAAGAAAGTTGATAACTTA
   ACTAAAGCTGAGAGGGGTGGCTTGTCTGAACCTTGACAAGGCCGGATTATTAAACGTCAGCTCGTGAAACCCGC
   CAAATCACAAAGCATGTTGCACAGATACTAGATTCCCGAATGAATACGAAATACGACGAGAACGATAAGCTGATT
   CGGGAAGTCAAAGTAATCACTTTAAAGTCAAAATTGGTGTGCGACTTCAGAAAGGATTTTCAATTCTATAAAGTT
   AGGGAGATAAATACTACCACCATGCGCACGACGCTTATCTTAATGCCGTGCTAGGGACCGCACTCATTAAAGAAA
45  TACCCGAAGCTAGAAAGTGAGTTTGTGTATGGTGATTACAAAGTTTATGACGTCCGTAAGATGATCGCGAAAAAGC
   GAACAGGAGATAGGCAAGGCTACAGCCAAATACTTCTTTATTCTAACATTATGAATTTCTTTAAGACGGAATC
   ACTCTGGCAAACCGGAGAGATACGCAAACGACCTTTAATTGAACCAATGGGGAGACAGGTGAAATCGTATGGGAT
   AAGGGAACTCGGACTTCGCGACGGTGAGAAAAGTTTTGCTCATTGCCCAAGTCAACATAGTAAAGAACTGAGGTG
   CAGACCGGAGGGTTTTCAAAGGAATCGATTCTTCCAAAAAGGAATAGTGATAAGCTCATCGCTCGTAAAAAGGAC
50  TGGGACCCGAAAAAGTACGGTGGCTTCGATAGCCCTACAGTTGCCTATTCTGTCTAGTAGTGCCAAAAGTTGAG
   AAGGGAAAAATCCAAGAACTGAAGTCAGTCAAAGAATTATTGGGGATAACGATTATGGAGCGCTCGTCTTTTGA
   AAGAACCCCATCGACTTCTTTGAGGCGAAAGGTTACAAGGAAGTAAAAAAGGATCTCATAATTAAGTACCAAAG
   TATAGTCTGTTTGAGTTAGAAAAATGGCCGAAACGGATGTTGGCTAGCGCCGGAGAGCTTCAAAGGGGACGAA
   CTCGCACTACGCTCTAAATACGTGAATTTCTGTATTTAGCGTCCATTACGAGAAGTTGAAAGGTTCACTGAA
55  GATAACGAACAGAAGCAACTTTTTGTTGAGCAGCACAAACATTATCTCGACGAAATCATAGAGCAAATTTGCGAA
   TTCAGTAAGAGAGTCATCTAGCTGATGCCAATCTGGACAAAGTATTAAGCGCATACAACAAGCACAGGGATAAAA
   CCCATACGTGAGCAGGCGGAAAAATATTATCCATTGTTTACTCTTACCAACCTCGGCGCTCCAGCCGCACTTCAAG
   TATTTTGACACAACGATAGATCGCAAACGATACACTTCTACCAAGGAGGTGCTAGACGCGACACTGATTACCAA
```

TCCATCACGGGATTATATGAAACTCGGATAGATTTGTACAGCTTGGGGGTGACGGATCCCCAAGAAGAAGAGG
AAAGTCTCGAGCGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGAC
AAGGCTGCAGGA (SEQ ID NO:3)

5 MDKKYSIGLAIGTNSVGWAVITDEYKVPSSKKFKVLGNTRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRR
KNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDFKFIQLVQTYNQLFEEENPINASGVDAKAILSARLSKSRRLENL
IAQLPGEKKNGFLGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDITYDDDLNLLAQIGDQYADFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKF
10 IKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPY
YVGPLARGNSRFAMWTRKSEETITPWNFEVVDKGASQSFIERMTNFDKNLPNEKVLPKHSLLEYEFTVYNELT
KVYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKI
IKDKDFLDNEENEDILEDIVLTTLTFEDREMIERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQS
GKTILDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQGSLSLHEHIANLAGSPAIAKKGILQTVKVVDELVKV
15 MGRHKPENIVIEAMARENQTTQKGQKNSRERMKRIEELGKSELKEHPVENTQLQNEKLYLYLQNGRDMYVD
QELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSEMNTRYDENDKLIREVKVITLKSCLVSDFRKDFQFYKV
REINNYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSEQEI GKATAKYFFYSNIMNFFKTEI
TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIKKTEVQTGGFSKESILPKRNSDKLIARKKD
20 WDPKYYGGFDSPTVAYSVLVAKVEKGSKKLKSVKELLGITIMERSSEFEKNPIDFLEAKGYKEVKKDLIIKLPK
YSLFELENGRKRMLASAGELQKGNELALPSKYVNFYLAHYEKLKGSPEDEQKQLFVEQHKHYLDEIEIEQISE
FSKRVLADANLDKVL SAYNKHDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDTLIHQ
SITGLYETRIDLSQLGGD (SEQ ID NO:4)

(single underline: HNH domain; double underline: RuvC domain)

25 In some embodiments, dCas9 corresponds to, or comprises in part or in whole, a Cas9
amino acid sequence having one or more mutations that inactivate the Cas9 nuclease activity.
For example, in some embodiments, a dCas9 domain comprises D10A and/or H820A
mutation.

30 dCas9 (D10A and H840A):

MDKKYSIGLAIGTNSVGWAVITDEYKVPSSKKFKVLGNTRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRR
KNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDFKFIQLVQTYNQLFEEENPINASGVDAKAILSARLSKSRRLENL
IAQLPGEKKNGFLGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDITYDDDLNLLAQIGDQYADFLAAKNLSDAI
35 LLSLILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKF
IKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPY
YVGPLARGNSRFAMWTRKSEETITPWNFEVVDKGASQSFIERMTNFDKNLPNEKVLPKHSLLEYEFTVYNELT
KVYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKI
IKDKDFLDNEENEDILEDIVLTTLTFEDREMIERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQS
40 GKTILDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQGSLSLHEHIANLAGSPAIAKKGILQTVKVVDELVKV
MGRHKPENIVIEAMARENQTTQKGQKNSRERMKRIEELGKSELKEHPVENTQLQNEKLYLYLQNGRDMYVD
QELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSEMNTRYDENDKLIREVKVITLKSCLVSDFRKDFQFYKV
REINNYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSEQEI GKATAKYFFYSNIMNFFKTEI
45 TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIKKTEVQTGGFSKESILPKRNSDKLIARKKD
WDPKYYGGFDSPTVAYSVLVAKVEKGSKKLKSVKELLGITIMERSSEFEKNPIDFLEAKGYKEVKKDLIIKLPK
YSLFELENGRKRMLASAGELQKGNELALPSKYVNFYLAHYEKLKGSPEDEQKQLFVEQHKHYLDEIEIEQISE
FSKRVLADANLDKVL SAYNKHDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDTLIHQ
SITGLYETRIDLSQLGGD (SEQ ID NO: 5)

50 (single underline: HNH domain; double underline: RuvC domain)

In other embodiments, dCas9 variants having mutations other than D10A and H820A
are provided, which *e.g.*, result in nuclease inactivated Cas9 (dCas9). Such mutations, by

way of example, include other amino acid substitutions at D10 and H820, or other substitutions within the nuclease domains of Cas9 (*e.g.*, substitutions in the HNH nuclease subdomain and/or the RuvC1 subdomain). In some embodiments, variants or homologues of dCas9 (*e.g.*, variants of SEQ ID NO: 5) are provided which are at least about 70% identical, at least about 80% identical, at least about 90% identical, at least about 95% identical, at least about 98% identical, at least about 99% identical, at least about 99.5% identical, or at least about 99.9% to SEQ ID NO:5. In some embodiments, variants of dCas9 (*e.g.*, variants of SEQ ID NO: 5) are provided having amino acid sequences which are shorter, or longer than SEQ ID NO: 5, by about 5 amino acids, by about 10 amino acids, by about 15 amino acids, by about 20 amino acids, by about 25 amino acids, by about 30 amino acids, by about 40 amino acids, by about 50 amino acids, by about 75 amino acids, by about 100 amino acids, or more.

In some embodiments, Cas9 fusion proteins as provided herein comprise the full-length amino acid of a Cas9 protein, *e.g.*, one of the sequences provided above. In other embodiments, however, fusion proteins as provided herein do not comprise a full-length Cas9 sequence, but only a fragment thereof. For example, in some embodiments, a Cas9 fusion protein provided herein comprises a Cas9 fragment, wherein the fragment binds crRNA and tracrRNA or sgRNA, but does not comprise a functional nuclease domain, *e.g.*, in that it comprises only a truncated version of a nuclease domain or no nuclease domain at all. Exemplary amino acid sequences of suitable Cas9 domains and Cas9 fragments are provided herein, and additional suitable sequences of Cas9 domains and fragments will be apparent to those of skill in the art.

In some embodiments, Cas9 refers to Cas9 from: *Corynebacterium ulcerans* (NCBI Refs: NC_015683.1, NC_017317.1); *Corynebacterium diphtheria* (NCBI Refs: NC_016782.1, NC_016786.1); *Spiroplasma syrphidicola* (NCBI Ref: NC_021284.1); *Prevotella intermedia* (NCBI Ref: NC_017861.1); *Spiroplasma taiwanense* (NCBI Ref: NC_021846.1); *Streptococcus iniae* (NCBI Ref: NC_021314.1); *Belliella baltica* (NCBI Ref: NC_018010.1); *Psychroflexus torquis* (NCBI Ref: NC_018721.1); *Streptococcus thermophilus* (NCBI Ref: YP_820832.1); *Listeria innocua* (NCBI Ref: NP_472073.1); *Campylobacter jejuni* (NCBI Ref: YP_002344900.1); or *Neisseria meningitidis* (NCBI Ref: YP_002342100.1).

The teaching herein with respect to the delivery of Cas9 or Cas9-like molecules as the payload in the ARMs comprising minimal-ARRDC1 is provided by way of a non-limiting example. Those skilled in art will recognize that the disclosure provides a means to load

other payloads including but not limited to other targeted endonucleases, including other RNA guided DNA-nucleases such as are known in the art that are in need of intracellular delivery.

The term “deaminase” refers to an enzyme that catalyzes a deamination reaction. In some embodiments, the deaminase is a cytidine deaminase, catalyzing the hydrolytic deamination of cytosine to uracil.

Cargo protein: The term “cargo protein”, as used herein, refers to a protein that may be incorporated in an ARMM, for example, into the liquid phase of the ARMM or into the lipid bilayer of an ARMM. The term “cargo protein to be delivered” refers to any protein that can be delivered via its association with or inclusion in an ARMM to a subject, organ, tissue, or cell. In some embodiments, the cargo protein is to be delivered to a target cell *in vitro*, *in vivo*, or *ex vivo*. In some embodiments, the cargo protein to be delivered is a biologically active agent, *i.e.*, it has activity in a cell, organ, tissue, and/or subject. For instance, a protein that, when administered to a subject, has a biological effect on that subject, is considered to be biologically active. In certain embodiments the cargo protein is a nuclease, deaminase, recombinase, or variant thereof (*e.g.*, a Cas9 protein or variant thereof). In certain embodiments, the nuclease may be a Cas9 nuclease, a TALE nuclease, a zinc finger nuclease, or any variant thereof. Nucleases, including Cas9 proteins and their variants, are described in more detail elsewhere herein. In some embodiments, the Cas9 protein or variant thereof is associated with a nucleic acid. For example, the cargo protein may be a Cas9 protein associated with a gRNA. In some embodiments, a cargo protein to be delivered is a therapeutic agent. As used herein, the term “therapeutic agent” refers to any agent that, when administered to a subject, has a beneficial effect. In some embodiments, the cargo protein to be delivered to a cell is a transcription factor, a tumor suppressor, a developmental regulator, a growth factor, a metastasis suppressor, a pro-apoptotic protein, a nuclease, a immunoglobulin or fragment thereof, a receptor, a T-cell receptor, a cytokine, an enzyme or a recombinase. In some embodiments, the protein to be delivered is p53, Rb (retinoblastoma protein), BRCA1, BRCA2, PTEN, APC, CD95, ST7, ST14, a BCL-2 family protein, a caspase, BRMS1, CRSP3, DRG1, KAI1, KISS1, NM23, a TIMP-family protein, a BMP-family growth factor, EGF, EPO, FGF, G-CSF, GM-CSF, a GDF-family growth factor, HGF, HDGF, IGF, PDGF, TPO, TGF- α , TGF- β , VEGF; a zinc finger nuclease, Cre recombinase, Dre recombinase, or FLP recombinase. In some embodiments, the cargo protein is associated with a small molecule. In some embodiments, the cargo protein to be delivered is a diagnostic agent. In some embodiments, the cargo protein to be delivered is a prophylactic

agent. In some embodiments, the cargo protein to be delivered is useful as an imaging agent. In some of these embodiments, the diagnostic or imaging agent is, and in others it is not, biologically active.

Conserved: As used herein, the term “conserved” refers to nucleotides or amino acid residues of a polynucleotide sequence or amino acid sequence, respectively, that are those that occur unaltered in the same position of two or more related sequences being compared. Nucleotides or amino acids that are relatively conserved are those that are conserved amongst more related sequences than nucleotides or amino acids appearing elsewhere in the sequences. In some embodiments, two or more sequences are said to be “completely conserved” if they are 100% identical to one another. In some embodiments, two or more sequences are said to be “highly conserved” if they are at least 70% identical, at least 80% identical, at least 90% identical, or at least 95% identical to one another. In some embodiments, two or more sequences are said to be “highly conserved” if they are about 70% identical, about 80% identical, about 90% identical, about 95% identical, about 98% identical, or about 99% identical to one another. In some embodiments, two or more sequences are said to be “conserved” if they are at least 60% identical, at least 70% identical, at least 80% identical, at least 90% identical, or at least 95% identical to one another. In some embodiments, two or more sequences are said to be “conserved” if they are about 30% identical, about 40% identical, about 50% identical, about 60% identical, about 70% identical, about 80% identical, about 90% identical, about 95% identical, about 98% identical, or about 99% identical to one another.

Engineered: The term “engineered,” as used herein, refers to a protein, nucleic acid, complex, substance, or entity that has been designed, produced, prepared, synthesized, and/or manufactured by a human. Accordingly, an engineered product is a product that does not occur in nature. In some embodiments, an engineered protein or nucleic acid is a protein or nucleic acid that has been designed to meet particular requirements or to have particular design features. For example, a Cas9 cargo protein may be engineered to associate with the minimal ARRDC1 by fusing one or more WW domains to the Cas9 protein to facilitate loading of the Cas9 cargo protein into an ARMM. As another example, a guide RNA (gRNA) may be engineered to target the delivery of a Cas9 cargo protein to a specific genomic sequence.

Expression: As used herein, “expression” of a nucleic acid sequence refers to one or more of the following events: (1) production of an RNA template from a DNA sequence (e.g., by transcription); (2) processing of an RNA transcript (e.g., by splicing, editing, 5' cap

formation, and/or 3' end processing); (3) translation of an RNA into a polypeptide or protein; and (4) post-translational modification of a polypeptide or protein.

Functional: As used herein, a “functional” biological molecule is a biological molecule in a form in which it exhibits a property and/or activity by which it is characterized
5 or useful.

Fusion protein: As used herein, a “fusion protein” includes a first protein moiety, e.g., an ARRC1 protein or fragment thereof, or a TSG101 protein or fragment thereof, having a peptide linkage with a second protein moiety, for example, a protein to be delivered to a target cell. In certain embodiments, the fusion protein is encoded by a single fusion
10 gene.

Gene: As used herein, the term “gene” has its meaning as understood in the art. It will be appreciated by those of ordinary skill in the art that the term “gene” may include gene regulatory sequences (e.g., promoters, enhancers, etc.) and/or intron sequences. It will further be appreciated that definitions of gene include references to nucleic acids that do not
15 encode proteins but rather encode functional RNA molecules, such as gRNAs, RNAi agents, ribozymes, tRNAs, rRNAs, etc. For the purpose of clarity, it should be noted that, as used in the present application, the term “gene” generally refers to a portion of a nucleic acid that encodes a protein; the term may optionally encompass regulatory sequences, as will be clear from context to those of ordinary skill in the art. This definition is not intended to exclude
20 application of the term “gene” to non-protein-coding expression units but rather to clarify that, in most cases, the term as used herein refers to a protein-coding nucleic acid.

Gene product or expression product: As used herein, the term “gene product” or “expression product” generally refers to an RNA transcribed from the gene (pre- and/or post-processing) or a polypeptide (pre- and/or post-modification) encoded by an RNA transcribed
25 from the gene.

Green fluorescent protein: As used herein, the term “green fluorescent protein” (GFP) refers to a protein originally isolated from the jellyfish *Aequorea victoria* that fluoresces green when exposed to blue light or a derivative of such a protein (e.g., an enhanced or wavelength-shifted version of the protein). The amino acid sequence of wild
30 type GFP is as follows:

MSKGEELFTG VVPILVELDG DVNGHKFSVS GEGEGDATYG
KLTLKFICTT GKLPVPWPTL VTTFSGYGVQC FSRYPDHMKQ
HDFFKSAMPE GYVQERTIFF KDDGNYKTRA EVKFEGDTLV
NRIELKGIDF KEDGNILGHK LEYNYNSHNV YIMADKQKNG

IKVNFKIRHN IEDGSVQLAD HYQQNTPIGD GPVLLPDNHY
LSTQSALSKD PNEKRDHMLV LEFVTAAGIT HGMDLEYK (SEQ ID NO:
35)

Proteins that are at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at
5 least 95%, at least 98%, or at least 99% identical are also considered to be green fluorescent
proteins.

Homology: As used herein, the term “homology” refers to the overall relatedness
between polymeric molecules, e.g., between nucleic acid molecules (e.g. DNA molecules
and/or RNA molecules) and/or between polypeptide molecules. In some embodiments,
10 polymeric molecules are considered to be “homologous” to one another if their sequences are
at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%,
at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%,
at least 95%, or at least 99% identical. In some embodiments, polymeric molecules are
considered to be “homologous” to one another if their sequences are at least 25%, at least
15 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least
65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or at
least 99% similar. The term “homologous” necessarily refers to a comparison between at
least two sequences (nucleotide sequences or amino acid sequences). In accordance with the
invention, two nucleotide sequences are considered to be homologous if the polypeptides
20 they encode are at least about 50% identical, at least about 60% identical, at least about 70%
identical, at least about 80% identical, or at least about 90% identical for at least one stretch
of at least about 20 amino acids. In some embodiments, homologous nucleotide sequences
are characterized by the ability to encode a stretch of at least 4–5 uniquely specified amino
acids. Both the identity and the approximate spacing of these amino acids relative to one
25 another must be considered for nucleotide sequences to be considered homologous. For
nucleotide sequences less than 60 nucleotides in length, homology is determined by the
ability to encode a stretch of at least 4-5 uniquely specified amino acids. In accordance with
the invention, two protein sequences are considered to be homologous if the proteins are at
least about 50% identical, at least about 60% identical, at least about 70% identical, at least
30 about 80% identical, or at least about 90% identical for at least one stretch of at least about 20
amino acids.

Identity: As used herein, the term “identity” refers to the overall relatedness between
polymeric molecules, e.g., between nucleic acid molecules (e.g. DNA molecules and/or RNA
molecules) and/or between polypeptide molecules. Calculation of the percent identity of two

nucleic acid sequences, for example, can be performed by aligning the two sequences for optimal comparison purposes (e.g., gaps can be introduced in one or both of a first and a second nucleic acid sequences for optimal alignment and non-identical sequences can be disregarded for comparison purposes). In certain embodiments, the length of a sequence aligned for comparison purposes is at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, or 100% of the length of the reference sequence. The nucleotides at corresponding nucleotide positions are then compared. When a position in the first sequence is occupied by the same nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position. The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, which needs to be introduced for optimal alignment of the two sequences. The comparison of sequences and determination of percent identity between two sequences can be accomplished using a mathematical algorithm. For example, the percent identity between two nucleotide sequences can be determined using methods such as those described in COMPUTATIONAL MOLECULAR BIOLOGY, Lesk, A. M., ed., Oxford University Press, New York, 1988; BIOCOMPUTING: INFORMATICS AND GENOME PROJECTS, Smith, D. W., ed., Academic Press, New York, 1993; SEQUENCE ANALYSIS IN MOLECULAR BIOLOGY, von Heinje, G., Academic Press, 1987; COMPUTER ANALYSIS OF SEQUENCE DATA, PART I, Griffin, A. M., and Griffin, H. G., eds., Humana Press, New Jersey, 1994; and SEQUENCE ANALYSIS PRIMER, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991; each of which is incorporated herein by reference. For example, the percent identity between two nucleotide sequences can be determined using the algorithm of Meyers and Miller (CABIOS, 1989, 4:11-17), which has been incorporated into the ALIGN program (version 2.0) using a PAM120 weight residue table, a gap length penalty of 12 and a gap penalty of 4. The percent identity between two nucleotide sequences can, alternatively, be determined using the GAP program in the GCG software package using an NWSgapdna.CMP matrix. Methods commonly employed to determine percent identity between sequences include, but are not limited to those disclosed in Carillo, H., and Lipman, D., SIAM J Applied Math., 48:1073 (1988); incorporated herein by reference. Techniques for determining identity are codified in publicly available computer programs. Exemplary computer software to determine homology between two sequences include, but are not limited to, GCG program package, Devereux, J., et al., *Nucleic Acids Research*, 12(1), 387 (1984)), BLASTP, BLASTN, and FASTA Atschul, S. F. et al., *J. Molec. Biol.*, 215, 403 (1990)).

Inhibit expression of a gene: As used herein, the phrase “inhibit expression of a gene” means to cause a reduction in the amount of an expression product of the gene. The expression product can be an RNA transcribed from the gene (e.g., an mRNA) or a polypeptide translated from an mRNA transcribed from the gene. Typically a reduction in the level of an mRNA results in a reduction in the level of a polypeptide translated therefrom. The level of gene expression may be determined using standard techniques for measuring mRNA and/or protein levels.

In vitro: As used herein, the term “in vitro” refers to events that occur in an artificial environment, e.g., in a test tube or reaction vessel, in cell culture, in a Petri dish, etc., rather than within an organism (e.g., animal, plant, or microbe).

In vivo: As used herein, the term “in vivo” refers to events that occur within an organism (e.g., animal, plant, or microbe).

Isolated: As used herein, the term “isolated” refers to a substance or entity that has been (1) separated from at least some of the components with which it was associated when initially produced (whether in nature or in an experimental setting), and/or (2) produced, prepared, and/or manufactured by the hand of man. Isolated substances and/or entities may be separated from at least about 10%, about 20%, about 30%, about 40%, about 50%, about 60%, about 70%, about 80%, about 90%, or more of the other components with which they were initially associated. In some embodiments, isolated substances are more than about 80%, about 85%, about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, about 99%, or more than about 99% pure. As used herein, a substance is “pure” if it is substantially free of other components.

Linker: The term “linker,” as used herein, refers to a chemical moiety linking two molecules or moieties, e.g., a minimal ARRDC1 protein and a target, such as a Cas9 nuclease. Typically, the linker is positioned between, or flanked by, two groups, molecules, or other moieties and connected to each one via a covalent bond, thus connecting the two. In some embodiments, the linker comprises an amino acid or a plurality of amino acids (e.g., a peptide or protein). In some embodiments, the linker is an organic molecule, group, polymer, or other chemical moiety. In some embodiments, the linker is a cleavable linker, e.g., the linker comprises a bond that can be cleaved upon exposure to, for example, UV light or a hydrolytic enzyme, such as a protease or esterase. In some embodiments, the linker is any stretch of amino acids having at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 15, at least 20, at least 25, at least 30, at least

40, at least 50, or more amino acids. In other embodiments, the linker is a chemical bond (e.g., a covalent bond).

microRNA (miRNA): As used herein, the term “microRNA” or “miRNA” refers to an RNAi agent that is approximately 21 nucleotides (nt) – 23 nt in length. miRNAs can range
5 between 18 nt – 26 nt in length. Typically, miRNAs are single-stranded. However, in some embodiments, miRNAs may be at least partially double-stranded. In certain embodiments, miRNAs may comprise an RNA duplex (referred to herein as a “duplex region”) and may optionally further comprises one to three single-stranded overhangs. In some embodiments, an RNAi agent comprises a duplex region ranging from 15 bp to 29 bp in length and
10 optionally further comprising one or two single-stranded overhangs. An miRNA may be formed from two RNA molecules that hybridize together, or may alternatively be generated from a single RNA molecule that includes a self-hybridizing portion. In general, free 5' ends of miRNA molecules have phosphate groups, and free 3' ends have hydroxyl groups. The duplex portion of an miRNA usually, but does not necessarily, comprise one or more bulges
15 consisting of one or more unpaired nucleotides. One strand of an miRNA includes a portion that hybridizes with a target RNA. In certain embodiments, one strand of the miRNA is not precisely complementary with a region of the target RNA, meaning that the miRNA hybridizes to the target RNA with one or more mismatches. In some embodiments, one strand of the miRNA is precisely complementary with a region of the target RNA, meaning
20 that the miRNA hybridizes to the target RNA with no mismatches. Typically, miRNAs are thought to mediate inhibition of gene expression by inhibiting translation of target transcripts. However, in some embodiments, miRNAs may mediate inhibition of gene expression by causing degradation of target transcripts.

The term “microvesicle,” as used herein, refers to a droplet of liquid surrounded by a
25 lipid bilayer. In some embodiments, a microvesicle has a diameter of about 10 nm to about 1000 nm. In some embodiments, a microvesicle has a diameter of at least about 10 nm, at least about 20 nm, at least about 30 nm, at least about 40 nm, at least about 50 nm, at least about 60 nm, at least about 70 nm, at least about 80 nm, at least about 90 nm, at least about 100 nm, at least about 125 nm, at least about 150 nm, at least about 175 nm, at least about
30 200 nm, at least about 250 nm, at least about 300 nm, at least about 400 nm, or at least about 500 nm. In some embodiments, a microvesicle has a diameter of less than about 1000 nm, less than about 900 nm, less than about 800 nm, less than about 700 nm, less than about 600 nm, less than about 500 nm, less than about 400 nm, less than about 300 nm, less than about 250 nm, less than about 200 nm, less than about 150 nm, less than about 100 nm, less than

about 90 nm, less than about 80 nm, less than about 70 nm, less than about 60 nm, or less than about 50 nm. The term microvesicle includes microvesicle shed from cells as well as synthetically produced microvesicles. Microvesicles shed from cells typically comprise the antigenic content of the cells from which they originate. Microvesicles shed from cells also typically comprise an asymmetric distribution of phospholipids, reflecting the phospholipid distribution of the cells from which they originate. In some embodiments, the inner membrane of microvesicles provided herein, e.g., of some ARMMs, comprises the majority of aminophospholipids, phosphatidylserine, and/or phosphatidylethanolamine within the lipid bilayer.

10 *Minimal ARRDC1*: As used herein, the term “minimal ARRDC1” refers to a ARRDC1 protein that is shorter than the full-length ARRDC1 protein, and comprises, at least, a portion of an arrestin domain, at least one PSAP (SEQ ID NO: 122) or PTAP (SEQ ID NO: 123) motif, and at least one PPXY motif. An exemplary minimal ARRDC1 is provided in SEQ ID NO: 1.

15 MGRVQLFEISLSHGRVVYSPGEPLAGTVRVRLGAPLPFRAIRVTCIGSCGVSNKANDT
AWVVEEGYFNSSLADKGSPLAGEHSFPFQFLLPATAPTSFEGPFGKIVHQQVRAAIHT
PRFSKDHKCSLVFYILSPLNLNSIPDIEQPNVASATKKFSYKLVKTGSVVLTASTDLRGY
VVGQALQLHADVENQSGKDTSPVVASLLQKVSYKAKRWIHDVRTIAEVEGAGVKA
WRRAQWHEQILVPALPQSALPGCSLIHIDYYLQVSLKAPEATVTLPVFIGNIAVNHAP
20 VSPRPGLGLPPGAPPLVV**PSAPPOEEAEPPEYPYEAPPSY**
(SEQ ID NO: 1).

Nucleic acid: As used herein, the term “nucleic acid,” in its broadest sense, refers to any compound and/or substance that is or can be incorporated into an oligonucleotide chain. In some embodiments, a nucleic acid is a compound and/or substance that is or can be incorporated into an oligonucleotide chain via a phosphodiester linkage. In some

25 embodiments, “nucleic acid” refers to individual nucleic acid residues (e.g., nucleotides and/or nucleosides). In some embodiments, “nucleic acid” refers to an oligonucleotide chain comprising individual nucleic acid residues. As used herein, the terms “oligonucleotide” and “polynucleotide” can be used interchangeably to refer to a polymer of nucleotides (e.g., a

30 string of at least two nucleotides). In some embodiments, “nucleic acid” encompasses RNA as well as single and/or double-stranded DNA. Furthermore, the terms “nucleic acid,” “DNA,” “RNA,” and/or similar terms include nucleic acid analogs, i.e., analogs having other than a phosphodiester backbone. For example, the so-called “peptide nucleic acids,” which are known in the art and have peptide bonds instead of phosphodiester bonds in the backbone,

are considered within the scope of the present invention. The term “nucleotide sequence encoding an amino acid sequence” includes all nucleotide sequences that are degenerate versions of each other and/or encode the same amino acid sequence. Nucleotide sequences that encode proteins and/or RNA may include introns. Nucleic acids can be purified from natural sources, produced using recombinant expression systems and optionally purified, chemically synthesized, etc. Where appropriate, e.g., in the case of chemically synthesized molecules, nucleic acids can comprise nucleoside analogs such as analogs having chemically modified bases or sugars, backbone modifications, etc. A nucleic acid sequence is presented in the 5' to 3' direction unless otherwise indicated. The term “nucleic acid segment” is used herein to refer to a nucleic acid sequence that is a portion of a longer nucleic acid sequence. In many embodiments, a nucleic acid segment comprises at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, or more residues. In some embodiments, a nucleic acid is or comprises natural nucleosides (e.g., adenosine, thymidine, guanosine, cytidine, uridine, deoxyadenosine, deoxythymidine, deoxyguanosine, and deoxycytidine); nucleoside analogs (e.g., 2-aminoadenosine, 2-thiothymidine, inosine, pyrrolo-pyrimidine, 3-methyl adenosine, 5-methylcytidine, 2-aminoadenosine, C5-bromouridine, C5-fluorouridine, C5-iodouridine, C5-propynyl-uridine, C5-propynyl-cytidine, C5-methylcytidine, 2-aminoadenosine, 7-deazaadenosine, 7-deazaguanosine, 8-oxoadenosine, 8-oxoguanosine, O(6)-methylguanine, and 2-thiocytidine); chemically modified bases; biologically modified bases (e.g., methylated bases); intercalated bases; modified sugars (e.g., 2'-fluororibose, ribose, 2'-deoxyribose, arabinose, and hexose); and/or modified phosphate groups (e.g., phosphorothioates and 5'-N-phosphoramidite linkages). In some embodiments, the present invention is specifically directed to “unmodified nucleic acids,” meaning nucleic acids (e.g., polynucleotides and residues, including nucleotides and/or nucleosides) that have not been chemically modified in order to facilitate or achieve delivery.

Protein: The term “protein” is used herein interchangeably with the terms polypeptide and peptide, and refers to a polypeptide (i.e., a string of at least two amino acids linked to one another by peptide bonds). Proteins may include moieties other than amino acids (e.g., may be glycoproteins) and/or may be otherwise processed or modified. Those of ordinary skill in the art will appreciate that a “protein” can be a complete polypeptide chain as produced by a cell (with or without a signal sequence), or can be a functional portion thereof. Those of ordinary skill will further appreciate that a protein can sometimes include more than one polypeptide chain, for example linked by one or more disulfide bonds or associated by other means. Polypeptides may contain L-amino acids, D-amino acids, or both

and may contain any of a variety of amino acid modifications or analogs known in the art. Useful modifications include, e.g., addition of a chemical entity such as a carbohydrate group, a phosphate group, a farnesyl group, an isofarnesyl group, a fatty acid group, an amide group, a terminal acetyl group, a linker for conjugation, functionalization, or other
 5 modification (e.g., alpha amidation), etc. In certain embodiments, the modifications of the peptide lead to a more stable peptide (e.g., greater half-life *in vivo*). These modifications may include cyclization of the peptide, the incorporation of D-amino acids, etc. None of the modifications should substantially interfere with the desired biological activity of the peptide. In certain embodiments, the modifications of the peptide lead to a more biologically active
 10 peptide. In some embodiments, polypeptides may comprise natural amino acids, non-natural amino acids, synthetic amino acids, amino acid analogs, and combinations thereof.

Recombinase: The term “recombinase,” as used herein, refers to a site-specific enzyme that mediates the recombination of DNA between recombinase recognition sequences, which results in the excision, integration, inversion, or exchange (e.g.,
 15 translocation) of DNA fragments between the recombinase recognition sequences. Recombinases can be classified into two distinct families: serine recombinases (e.g., resolvases and invertases) and tyrosine recombinases (e.g., integrases). Examples of serine recombinases include, without limitation, Hin, Gin, Tn3, β -six, CinH, ParA, $\gamma\delta$, Bxb1, ϕ C31, TP901, TG1, ϕ BT1, R4, ϕ RV1, ϕ FC1, MR11, A118, U153, and gp29. Examples of tyrosine
 20 recombinases include, without limitation, Cre, FLP, R, Lambda, HK101, HK022, and pSAM2. The serine and tyrosine recombinase names stem from the conserved nucleophilic amino acid residue that the recombinase uses to attack the DNA and which becomes covalently linked to the DNA during strand exchange. Recombinases have numerous applications, including the creation of gene knockouts/knock-ins and gene therapy
 25 applications. See, e.g., Brown *et al.*, “Serine recombinases as tools for genome engineering.” *Methods*. 2011;53(4):372-9; Hirano *et al.*, “Site-specific recombinases as tools for heterologous gene integration.” *Appl. Microbiol. Biotechnol.* 2011; 92(2):227-39; Chavez and Calos, “Therapeutic applications of the Φ C31 integrase system.” *Curr. Gene Ther.* 2011;11(5):375-81; Turan and Bode, “Site-specific recombinases: from tag-and-target- to
 30 tag-and-exchange-based genomic modifications.” *FASEB J.* 2011; 25(12):4088-107; Venken and Bellen, “Genome-wide manipulations of *Drosophila melanogaster* with transposons, FLP recombinase, and Φ C31 integrase.” *Methods Mol. Biol.* 2012; 859:203-28; Murphy, “Phage recombinases and their applications.” *Adv. Virus Res.* 2012; 83:367-414; Zhang *et al.*, “Conditional gene manipulation: Cre-ating a new biological era.” *J. Zhejiang Univ. Sci. B.*

2012; 13(7):511-24; Karpenshif and Bernstein, "From yeast to mammals: recent advances in genetic control of homologous recombination." DNA Repair (Amst). 2012; 1;11(10):781-8; the entire contents of each are hereby incorporated by reference in their entirety. The recombinases provided herein are not meant to be exclusive examples of recombinases that can be used in embodiments of the invention. The methods and compositions of the invention can be expanded by mining databases for new orthogonal recombinases or designing synthetic recombinases with defined DNA specificities (See, *e.g.*, Groth *et al.*, "Phage integrases: biology and applications." J. Mol. Biol. 2004; 335, 667-678; Gordley *et al.*, "Synthesis of programmable integrases." Proc. Natl. Acad. Sci. U S A. 2009; 106, 5053-5058; the entire contents of each are hereby incorporated by reference in their entirety). Other examples of recombinases that are useful in the methods and compositions described herein are known to those of skill in the art, and any new recombinase that is discovered or generated is expected to be able to be used in the different embodiments of the invention. In some embodiments, a recombinase (or catalytic domain thereof) is fused to a Cas9 protein (*e.g.*, dCas9).

Recombine: The term "recombine" or "recombination," in the context of a nucleic acid modification (*e.g.*, a genomic modification), is used to refer to the process by which two or more nucleic acid molecules, or two or more regions of a single nucleic acid molecule, are modified by the action of a recombinase protein. Recombination can result in, inter alia, the insertion, inversion, excision, or translocation of a nucleic acid sequence, *e.g.*, in or between one or more nucleic acid molecules.

Reprogramming factor: As used herein, the term "reprogramming factor" refers to a factor that, alone or in combination with other factors, can change the state of a cell from a somatic, differentiated state into a pluripotent stem cell state. Non-limiting examples of reprogramming factors include a protein (*e.g.*, a transcription factor), a peptide, a nucleic acid, or a small molecule. Known reprogramming factors that are useful for cell reprogramming include, but are not limited to Oct4, Sox2, Klf4, and c-myc. Similarly, a programming factor may be used to modulate cell differentiation or de-differentiation, for example, to facilitate or induce cell differentiation towards a desired lineage.

RNA interference (RNAi): As used herein, the term "RNA interference" or "RNAi" refers to sequence-specific inhibition of gene expression and/or reduction in target RNA levels mediated by an RNA, which RNA comprises a portion that is substantially complementary to a target RNA. Typically, at least part of the substantially complementary portion is within the double stranded region of the RNA. In some embodiments, RNAi can

occur via selective intracellular degradation of RNA. In some embodiments, RNAi can occur by translational repression.

RNAi agent: As used herein, the term “RNAi agent” or “RNAi” refers to an RNA, optionally including one or more nucleotide analogs or modifications, having a structure characteristic of molecules that can mediate inhibition of gene expression through an RNAi mechanism. In some embodiments, RNAi agents mediate inhibition of gene expression by causing degradation of target transcripts. In some embodiments, RNAi agents mediate inhibition of gene expression by inhibiting translation of target transcripts. Generally, an RNAi agent includes a portion that is substantially complementary to a target RNA. In some embodiments, RNAi agents are at least partly double-stranded. In some embodiments, RNAi agents are single-stranded. In some embodiments, exemplary RNAi agents can include siRNA, shRNA, and/or miRNA. In some embodiments, RNAi agents may be composed entirely of natural RNA nucleotides (i.e., adenine, guanine, cytosine, and uracil). In some embodiments, RNAi agents may include one or more non-natural RNA nucleotides (e.g., nucleotide analogs, DNA nucleotides, *etc.*). Inclusion of non-natural RNA nucleic acid residues may be used to make the RNAi agent more resistant to cellular degradation than RNA. In some embodiments, the term “RNAi agent” may refer to any RNA, RNA derivative, and/or nucleic acid encoding an RNA that induces an RNAi effect (e.g., degradation of target RNA and/or inhibition of translation). In some embodiments, an RNAi agent may comprise a blunt-ended (i.e., without overhangs) dsRNA that can act as a Dicer substrate. For example, such an RNAi agent may comprise a blunt-ended dsRNA which is $\geq$ 25 base pairs length, which may optionally be chemically modified to abrogate an immune response.

RNAi-inducing agent: As used herein, the term “RNAi-inducing agent” encompasses any entity that delivers, regulates, and/or modifies the activity of an RNAi agent. In some embodiments, RNAi-inducing agents may include vectors (other than naturally occurring molecules not modified by the hand of man) whose presence within a cell results in RNAi and leads to reduced expression of a transcript to which the RNAi-inducing agent is targeted. In some embodiments, RNAi-inducing agents are RNAi-inducing vectors. In some embodiments, RNAi-inducing agents are compositions comprising RNAi agents and one or more pharmaceutically acceptable excipients and/or carriers. In some embodiments, an RNAi-inducing agent is an “RNAi-inducing vector,” which refers to a vector whose presence within a cell results in production of one or more RNAs that self-hybridize or hybridize to each other to form an RNAi agent (e.g. siRNA, shRNA, and/or miRNA). In various

embodiments, this term encompasses plasmids, e.g., DNA vectors (whose sequence may comprise sequence elements derived from a virus), or viruses (other than naturally occurring viruses or plasmids that have not been modified by the hand of man), whose presence within a cell results in production of one or more RNAs that self-hybridize or hybridize to each other to form an RNAi agent. In general, the vector comprises a nucleic acid operably linked to expression signal(s) so that one or more RNAs that hybridize or self-hybridize to form an RNAi agent are transcribed when the vector is present within a cell. Thus the vector provides a template for intracellular synthesis of the RNA or RNAs or precursors thereof. For purposes of inducing RNAi, presence of a viral genome in a cell (e.g., following fusion of the viral envelope with the cell membrane) is considered sufficient to constitute presence of the virus within the cell. In addition, for purposes of inducing RNAi, a vector is considered to be present within a cell if it is introduced into the cell, enters the cell, or is inherited from a parental cell, regardless of whether it is subsequently modified or processed within the cell. An RNAi-inducing vector is considered to be targeted to a transcript if presence of the vector within a cell results in production of one or more RNAs that hybridize to each other or self-hybridize to form an RNAi agent that is targeted to the transcript, i.e., if presence of the vector within a cell results in production of one or more RNAi agents targeted to the transcript.

RNA-programmable nuclease: The terms “RNA-programmable nuclease” and “RNA-guided nuclease” are used interchangeably herein and refer to a nuclease that forms a complex with (e.g., binds or associates with) one or more RNA molecule that is not a target for cleavage. In some embodiments, an RNA-programmable nuclease, when in a complex with an RNA, may be referred to as a nuclease:RNA complex. RNA-programmable nucleases include Cas9 nucleases. Typically, the bound RNA(s) is referred to as a guide RNA (gRNA). gRNAs can exist as a complex of two or more RNAs, or as a single RNA molecule. gRNAs that exist as a single RNA molecule may be referred to as single-guide RNAs (sgRNAs), though “gRNA” is used interchangeably to refer to guide RNAs that exist as either single molecules or as two or more molecules. Typically, gRNAs that exist as single RNA species comprise two domains: (1) a domain that shares homology to a target nucleic acid (e.g., and directs binding of a Cas9 complex to the target); and (2) a domain that binds an endonuclease enzyme (e.g., Cas9 protein). The gRNA comprises a nucleotide sequence that complements a target site, which mediates binding of the nuclease:RNA complex to said target site and providing the sequence specificity of the nuclease:RNA complex.

Short, interfering RNA (siRNA): As used herein, the term “short, interfering RNA” or “siRNA” refers to an RNAi agent comprising an RNA duplex (referred to herein as a “duplex region”) that is approximately 19 base pairs (bp) in length and optionally further comprises one to three single-stranded overhangs. In some embodiments, an RNAi agent comprises a duplex region ranging from 15 bp to 29 bp in length and optionally further comprising one or two single-stranded overhangs. An siRNA may be formed from two RNA molecules that hybridize together, or may alternatively be generated from a single RNA molecule that includes a self-hybridizing portion. In general, free 5'-ends of siRNA molecules have phosphate groups, and free 3'-ends have hydroxyl groups. The duplex portion of an siRNA may, but typically does not, comprise one or more bulges consisting of one or more unpaired nucleotides. One strand of an siRNA includes a portion that hybridizes with a target transcript. In certain embodiments, one strand of the siRNA is precisely complementary with a region of the target transcript, meaning that the siRNA hybridizes to the target transcript without a single mismatch. In some embodiments, one or more mismatches between the siRNA and the targeted portion of the target transcript may exist. In some embodiments in which perfect complementarity is not achieved, any mismatches are generally located at or near the siRNA termini. In some embodiments, siRNAs mediate inhibition of gene expression by causing degradation of target transcripts.

Short hairpin RNA (shRNA): As used herein, the term “short hairpin RNA” or “shRNA” refers to an RNAi agent comprising an RNA having at least two complementary portions hybridized or capable of hybridizing to form a double-stranded (duplex) structure sufficiently long to mediate RNAi (typically at least approximately 19 bp in length), and at least one single-stranded portion, typically ranging between approximately 1 nucleotide (nt) and approximately 10 nt in length that forms a loop. In some embodiments, an shRNA comprises a duplex portion ranging from 15 bp to 29 bp in length and at least one single-stranded portion, typically ranging between approximately 1 nt and approximately 10 nt in length that forms a loop. The duplex portion may, but typically does not, comprise one or more bulges consisting of one or more unpaired nucleotides. In some embodiments, siRNAs mediate inhibition of gene expression by causing degradation of target transcripts. shRNAs are thought to be processed into siRNAs by the conserved cellular RNAi machinery. Thus, shRNAs may be precursors of siRNAs. Regardless, siRNAs in general are capable of inhibiting expression of a target RNA, similar to siRNAs.

Small molecule: In general, a “small molecule” refers to a substantially non-peptidic, non-oligomeric organic compound either prepared in the laboratory or found in nature. Small

molecules, as used herein, can refer to compounds that are “natural product-like,” however, the term “small molecule” is not limited to “natural product-like” compounds. Rather, a small molecule is typically characterized in that it contains several carbon-carbon bonds, and has a molecular weight of less than 2000 g/mol, less than 1500 g/mol, less than 1250 g/mol, less than 1000 g/mol, less than 750 g/mol, less than 500 g/mol, or less than 250 g/mol, although this characterization is not intended to be limiting for the purposes of the present invention. In certain other embodiments, natural-product-like small molecules are utilized.

Similarity: As used herein, the term “similarity” refers to the overall relatedness between polymeric molecules, e.g., between nucleic acid molecules (e.g. DNA molecules and/or RNA molecules) and/or between polypeptide molecules. Calculation of percent similarity of polymeric molecules to one another can be performed in the same manner as a calculation of percent identity, except that calculation of percent similarity takes into account conservative substitutions as is understood in the art.

Subject: As used herein, the term “subject” or “patient” refers to any organism to which a composition in accordance with the invention may be administered, e.g., for experimental, diagnostic, prophylactic, and/or therapeutic purposes. Typical subjects include animals (e.g., mammals, such as mice, rats, rabbits, non-human primates, and humans) and/or plants.

Therapeutically effective amount: As used herein, the term “therapeutically effective amount” means an amount of an agent to be delivered (e.g., nucleic acid, drug, therapeutic agent, diagnostic agent, prophylactic agent, *etc.*) that is sufficient, when administered to a subject suffering from or susceptible to a disease, disorder, and/or condition, to treat, improve symptoms of, diagnose, prevent, and/or delay the onset of the disease, disorder, and/or condition.

Transcription factor: As used herein, the term “transcription factor” refers to a DNA-binding protein that regulates transcription of DNA into RNA, for example, by activation or repression of transcription. Some transcription factors effect regulation of transcription alone, while others act in concert with other proteins. Some transcription factor can both activate and repress transcription under certain conditions. In general, transcription factors bind a specific target sequence or sequences highly similar to a specific consensus sequence in a regulatory region of a target gene. Transcription factors may regulate transcription of a target gene alone or in a complex with other molecules. Examples of transcription factors include, but are not limited to, Sp1, NF1, CCAAT, GATA, HNF, PIT-1, MyoD, Myf5, Hox, Winged

Helix, SREBP, p53, CREB, AP-1, Mef2, STAT, R-SMAD, NF- κ B, Notch, TUBBY, and NFAT.

Treating: As used herein, the term “treating” refers to partially or completely alleviating, ameliorating, improving, relieving, delaying onset of, inhibiting progression of, reducing severity of, and/or reducing incidence of one or more symptoms or features of a particular disease, disorder, and/or condition. For example, “treating” cancer may refer to inhibiting survival, growth, and/or spread of a tumor. Treatment may be administered to a subject who does not exhibit signs of a disease, disorder, and/or condition and/or to a subject who exhibits only early signs of a disease, disorder, and/or condition for the purpose of decreasing the risk of developing pathology associated with the disease, disorder, and/or condition.

Vector: As used herein, “vector” refers to a nucleic acid molecule which can transport another nucleic acid to which it has been linked. In some embodiment, vectors can achieve extra-chromosomal replication and/or expression of nucleic acids to which they are linked in a host cell such as a eukaryotic and/or prokaryotic cell. Vectors capable of directing the expression of operatively linked genes are referred to herein as “expression vectors.”

WW Domain: The term “WW domain,” as described herein, is a protein domain having two basic residues at the C-terminus that mediates protein-protein interactions with short proline-rich or proline-containing motifs. The WW domain possessing the two basic C-terminal amino acid residues may have the ability to associate with short proline-rich or proline-containing motifs (*i.e.*, a PPXY motif). WW domains bind a variety of distinct peptide ligands including motifs with core proline-rich sequences, such as PPXY, which is found in AARDC1. A WW domain may be a 30-40 amino acid protein interaction domain with two signature tryptophan residues spaced by 20-22 amino acids. The three-dimensional structure of WW domains shows that they generally fold into a three-stranded, antiparallel β sheet with two ligand-binding grooves.

WW domains are found in many eukaryotes and are present in approximately 50 human proteins (Bork, P. & Sudol, M. The WW domain: a signaling site in dystrophin? Trends Biochem Sci 19, 531-533 (1994)). WW domains may be present together with several other interaction domains, including membrane targeting domains, such as C2 in the NEDD4 family proteins, the phosphotyrosine-binding (PTB) domain in FE65 protein, FF domains in CA150 and FBPII, and pleckstrin homology (PH) domains in PLEKHA5. WW domains are also linked to a variety of catalytic domains, including HECT E3 protein-ubiquitin ligase

domains in NEDD4 family proteins, rotomerase or peptidyl prolyl isomerase domains in Pin1, and Rho GAP domains in ArhGAP9 and ArhGAP12.

In the instant disclosure, the WW domain may be a WW domain that naturally possesses two basic amino acids at the C-terminus, for example a WW domain or WW domain variant may be from the human ubiquitin ligase WWP1, WWP2, Nedd4-1, Nedd4-2, Smurf1, Smurf2, ITCH, NEDL1, or NEDL2. Exemplary amino acid sequences of WW domain containing proteins (WW domains underlined) are listed below. It should be appreciated that any of the WW domains or WW domain variants of the exemplary proteins may be used in the invention, described herein, and are not meant to be limiting.

Human WWP1 amino acid sequence (uniprot.org/uniprot/Q9H0M0). The four underlined WW domains correspond to amino acids 349 – 382 (WW1), 381 – 414 (WW2), 456 – 489 (WW3), and 496 – 529 (WW4).

	MATASPRSDT	SNNHSGRLQL	QVTVSSAKLK	RKKNWFGTAI	YTEVVVDGEI	50
	TKTAKSSSSS	NPKWDEQLTV	NVTPQTTFLEF	QVWSHRTLKA	DALLGKATID	100
15	LKQALLIHNH	KLERVKEQLK	LSLENKNGIA	QTGELTVVLD	GLVIEQENIT	150
	NCSSSPTIEI	QENGDALHEN	GEPSARTTAR	LAVEGTNGID	NHVPTSTLVQ	200
	NSCCSYVUNG	DNTPSSPSQV	AARPKNTPAP	KPLASEPADD	TVNGESSSFA	250
	PTDNASVTGT	PVVSEENALS	PNCTSTTVED	PPVQEILTSS	ENNECIPSTS	300
	AELESEARSI	LEPDTSNSRS	SSAFEAAKSR	QPDGCMDFVR	QQSGNANTET	350
20	<u>LPSGWEQRKD</u>	<u>PHGRTYYVDH</u>	<u>NTRTTTWERP</u>	<u>QPLPPGWERR</u>	<u>VDDRRRVYYV</u>	400
	<u>DHNTRTTTWQ</u>	<u>RPTMESVRNF</u>	<u>EQWQSQRNQL</u>	<u>QCAMQQFNQR</u>	<u>YLYSASMLAA</u>	450
	<u>ENDPYGPLPP</u>	<u>GWEKRVDSTD</u>	<u>RVYFVNHNTK</u>	<u>TTQWEDPRTQ</u>	<u>GLQNEEPLPE</u>	500
	<u>GWEIRYTRREG</u>	<u>VRYFVDHNTR</u>	<u>TTTFKDPRNG</u>	<u>KSSVTKGGPQ</u>	<u>IAYERGFRWK</u>	550
	LAHFRYLCQS	NALPSHVKN	VSRQTLFEDS	FQIMALKPY	DLRRRLVYIF	600
25	RGEGLDYGG	LAREWFFLLS	HEVLNPMYCL	FEYAGKNNYC	LQINPASTIN	650
	PDHLSYFCFI	GRFIAMALFH	GKFIDTGFSL	PFYKRMLSKK	LTIKDLESID	700
	TEFYNSLIWI	RDNNIEECGL	EMYFSVDEMI	LGKVTSHDLK	LGGSNILVTE	750
	ENKDEYIGLM	TEWRFSRGVQ	EQTKAFLDGF	NEVVPQLWLQ	YFDEKELEV	800
	LCGMQEVDLA	DWQRNTVYRH	YTRNSKQIIW	FWQFVKETDN	EVRMRLQFV	850
30	TGTCRLPLGG	FAELMGSSNGP	QKFCIEKVGK	DTWLPRSHTC	FNRLDLPPYK	900
	SYEQLKEKLL	FAIETEGERG	QE	(SEQ ID NO: 6)		922

WW1 (349-382):

ETLPSGWEQRKDPHGRYTYVDHNTRTTTWERPQP (SEQ ID NO: 36).

35

WW2 (381-414):

QPLPPGWERRVDDRRRVYYVDHNTRTTTWQRPTM (SEQ ID NO: 37).

WW3 (456-489):

40 ENDPYGPLPPGWEKRVDSTDRVYFVNHNTKTTQWEDPRT (SEQ ID NO: 38).

WW4 (496-529):

EPLPEGWEIRYTRREGVRYFVDHNTRTTTFKDPRN (SEQ ID NO: 39).

Human WWP2 amino acid sequence (uniprot.org/uniprot/ O00308). The four underlined WW domains correspond to amino acids 300 – 333 (WW1), 330 – 363 (WW2), 405 – 437 (WW3), and 444 – 547 (WW4).

	MASASSSRAG	VALPFERSQL	TLKVVSAPK	VHNRQPRINS	YVEVAVDGLP	50
5	SETKKTGKRI	GSSELLWNEI	IILNVTAQSH	LDLKVWSCHT	LRNELLGTAS	100
	VNLSNVLKNN	GGKMNMQLT	LNLTENKGS	VVSGGELTIF	LDGPTVDLGN	150
	VPNGSALTDG	SQLPSRDSSG	TAVAPENRHQ	PPSTNCFGGR	SRTHRHSGAS	200
	ARTTFPATGEQ	SPGARSRRHQ	PVKNSGHSGL	ANGTVNDEPT	TATDPEEPSV	250
	VGVTSPPAAP	LSVTNPNTT	SLPAPATPAE	GEEPSTSGTQ	QLPAAAQAPD	300
10	<u>ALPAGWEQRE</u>	<u>LPNGRVYYVD</u>	<u>HNTKTTTWER</u>	<u>PLPPGWEKRT</u>	<u>DPRGRFYYVD</u>	350
	<u>HNRTTTTWR</u>	<u>PTAEYVRNVE</u>	<u>QWQSQRNQLQ</u>	<u>GAMQHFSQRF</u>	<u>LYQSSSASTD</u>	400
	<u>HDPLGPLPFG</u>	<u>WEKRQDNGRV</u>	<u>YYVNHNTRTT</u>	<u>QWEDPRTQGM</u>	<u>IQEPALPPGW</u>	450
	<u>EMKYTSEGVR</u>	<u>YFVDHNTRTT</u>	<u>TFKDPKPGFE</u>	<u>SGTKQGSFGA</u>	<u>YDRSFRWKYH</u>	500
	<u>QFRFLCHSNA</u>	<u>LPSHVKISVS</u>	<u>RQTLFEDSFQ</u>	<u>QIMNMKPYDL</u>	<u>RRRLYIIMRG</u>	550
15	EEGLDYGGA	REWFLLSHE	VLNPMYCLFE	YAGKNNYCLQ	INPASSINPD	600
	HLTYFRFIGR	FIAMALYHGK	FIDTGTTLFP	YKRLNKRPT	LKDLESIDPE	650
	FYNSIVWIK	NNLEECGLEL	YFIQDMEILG	KVTHELKEG	GESIRVTEEN	700
	KEEYIMLLTD	WRFTRGVEEQ	TKAFLDGFNE	VAPLEWLRYF	DEKELEMLC	750
	GMQEIDMSDW	QKSTIYRHYT	KNSKQIQFW	QVVKEMDNEK	RIRLLQFVTG	800
20	TCRLPVGGFA	ELIGSNGPQK	FCIDKVGKET	WLPRSHTCFN	RDLPPYKSY	850
	EQLREKLLYA	IEETEGFGQE	(SEQ ID NO: 7)			870

WW1 (300-333):

DALPAGWEQRELPNGRVYYVDHNTKTTTWERPLP (SEQ ID NO: 40).

25

WW2 (330-363):

PLPPGWEKRT DPRGRFYYVDHNTRTTTWQRPTA (SEQ ID NO: 41).

WW3 (405-437):

30 HDPLGPLPPGWEKRQDNGRVYYVNHNTRTTQWEDPRT (SEQ ID NO: 42).

WW4 (444-477):

PALPPGWEMKYTSEGVRYFVDHNTRTTTFKDPRP (SEQ ID NO: 43).

35 Human Nedd4-1 amino acid sequence (uniprot.org/uniprot/ P46934). The four underlined WW domains correspond to amino acids 610 – 643 (WW1), 767 – 800 (WW2), 840 – 873 (WW3), and 892 – 925 (WW4).

	MAQSLRLHFA	ARRSNTYPLS	ETSGDDLDSH	VHMCFKRPTR	ISTSNVVQMK	50
	LTPRQTALAP	LIKENVQSQE	RSSVPSSENV	NKKSSCLQIS	LQPTRYSGYL	100
40	QSSNVLADSD	DASFTCILKD	GIYSSAVVDN	ELNAVNDGHL	VSSPAICSGS	150
	LSNFSTSDNG	SYSSNGSDFG	SCASITSGGS	YTNSVISDSS	SYTFPPSDDT	200
	FLGGNLPDS	TSNRSVPNRN	TTPCEIFSR	TSTDFFVQDD	LEHGLEIMKL	250
	PVSRNTKIPL	KRYSSLVIFP	RSPSTTRPTS	PTSLCTLISK	GSYQTSQFI	300
	ISPSEIAHNE	DGTSAGFLS	TAVNGLRLSK	TICTPGEVRD	IRPLHRKGS	350
45	QKIVLSNNT	PRQTVCEKSS	EGYSCVSVHF	TQRKAATLDC	ETNGDCKPE	400

	MSEIKLNSDS	EYIKLMHRTS	ACLPSSQNVD	CQININGELE	RPHSQMNKNH	450
	GILRRSISLG	GAYPNISCLS	SLKHNCCKGG	PSQLLIKFFAS	GNEGKVDNLS	500
	RDSNRDCTNE	LSNSCKTRDD	FLGQVDVPLY	PLPTENPRLE	RPYTFKDFVL	550
	HPRSHKSRVK	GYLRLKMTYL	PKTSGSEDDN	AEQAELEPG	WVVDQPDAA	600
5	<u>CHLQQQQEPS</u>	<u>PLPPGWEEERQ</u>	<u>DILGRYYVN</u>	<u>HESRRTQWKR</u>	<u>PTPDNLTD</u>	650
	ENGNIQLQAQ	RAFTTRRQIS	EETESVDNRE	SSENWEIIRE	DEATMYSNQA	700
	FPSPPPSSNL	DVPTHLAEEL	NARLTIFGNS	AVSQPASSSN	HSSRRGSLQA	750
	YTFEEQPTLP	VLLPTSSGLP	PGWEEKQDER	GRSYYVDHNS	RTTTWTKPTV	800
	QATVETSQLT	SSQSSAGPQS	QASTSDSGQQ	VTQPSEIEQG	FLPKGWEVRH	850
10	<u>APNGRPFFID</u>	<u>HNTKTTTWED</u>	<u>PRLKIPAHLR</u>	<u>GKTSLDTSND</u>	<u>LGPLPPGWEE</u>	900
	<u>RTHTDGRIFY</u>	<u>INHNIKRTQW</u>	<u>EDPRLENVAI</u>	<u>TGPAVPYSRD</u>	<u>YKRKYEFFRR</u>	950
	KLKKQNDIPN	KFEMKLRRAT	VLEDSEYRRIM	GVKRADFLKA	RLWIEFDGEK	1000
	GLDYGGVARE	WFFLISKEMF	NPYYGLFEYS	ATDNYTLQIN	PNSGLCNEDH	1050
	LSYFKFIGRV	AGMAVYHGKL	LDGFFIRPFY	KMMLHKPITL	HDMSVDSEY	1100
15	<u>YNSLRWILEN</u>	<u>DPTLRLRFI</u>	<u>IDDEELFGQTH</u>	<u>QHELKNGGSE</u>	<u>IVVTNKNKKE</u>	1150
	<u>YIYLVIQWRF</u>	<u>VNRIQKQMAA</u>	<u>FKEGFFELIP</u>	<u>QDLIKIFDEN</u>	<u>ELELLMCGLG</u>	1200
	DVDVNDWREH	TKYKNGYSAN	HQVIQWFKA	VLMDSEKRI	RLQFVTGTS	1250
	KVPMNGFAEL	YGSNGPQSFT	VEQWGTPEKL	PRAHTCFNRL	DLPPYESFEE	1300
	LWDKLQMAIE	NTQCFDGV	(SEQ ID NO: 8)			1319
20						

WW1(610-643):

SPLPPGWEEERQDILGRYYVNHESRRTQWKRPTP (SEQ ID NO: 44).

WW2 (767-800):

25 SGLPPGWEEKQDERGRSYYVDHNSRTTTWTKPTV (SEQ ID NO: 45).

WW3 (840-873):

GFLPKGWEVRHAPNGRPFFIDHNTKTTTWEDPRL (SEQ ID NO: 46).

30 WW4 (892-925):

GPLPPGWEEERTHTDGRIFYINHNIKRTQWEDPRL (SEQ ID NO: 47).

Human Nedd4-2 amino acid sequence (>gil21361472|reflNP_056092.2| E3 ubiquitin-protein ligase NEDD4-like isoform 3 [*Homo sapiens*]). The four underlined WW domains
 35 correspond to amino acids 198 – 224 (WW1), 368 – 396 (WW2), 480 – 510 (WW3), and 531 – 561 (WW4).

MATGLGEPVYGLSEDEGESRILRVKVVSGIDLAKKDIFGASDPYVKLSLYVADENRE
 LALVQTKTIKKTLPKWNEEFYFRVNPSNHRLLEVFEDENRLTRDDFLGQVDVPLSH
 LPTEDPTMERPYTFKDFLLRPRSHKSRVKGFLRLKMA YMPKNGGQDEENSQDRDD
 40 MEHGWEVVDSDNSASQHQEELPPPPLPPGWEEKVDNLGRYYVNHNNRTTQWHRP
 SLMDVSSESDNNIRQINQEAHRRFRSRRHISEDLEPEPSEGGDVPEPWETISEEVNIA
 GDSLGLALPPPPASPGSRTSPQELSEELSRRLQITPDSNGEQFSSLIQREPSSRLRSCSVT

DAVAEQGHLPPPSVAYVHTTPGLPSGWEERKDAKGRTYVNHNNRTTTTWTRPIMQ
 LAEDGASGSATNSNNHLIEPQIRRPRLSSPTVTLSAPLEGAKDSPVRRRAVKDTLSNP
 QSPQSPYNSPKPQHKVTQSFLPPGWEMRIAPNGRPFFIDHNTKTTTTWEDPRLKFPVH
 MRSKTSLNPNDLGPLPPGWEERIHLDGRTFYIDHNSKITQWEDPRLQNPAITGPAVPY
 5 SREFKQKYDYFRKKLKKPADIPNRFEMKLHRNNIFEESYRRIMSVKRPDVLKARLWI
 EFESEKGLDYGGVAREWFFLLSKEMFNPHYGLFEYSATDNYTLQINPNSGLCNEDHL
 SYFTFIGRVAGLAVFHGKLLDGFFIRPFYKMMLGKQITLNDMESVDSEYYNSLKWIL
 ENDPTELDLMFCIDEENFGQTYQVDLKPNGSEIMVTNENKREYIDLVIQWRFVNRVQ
 KQMNAFLEGFTELLPIDLIKIFDENELELLMCGLGDVDVNDWRQHHSYKNGYCPNHP
 10 VIQWFWKAVLLMDAEKRIRLLQFVTGTSRVPMNGFAELYGSNGPQLFTIEQWGSPE
 KLPRAHTCFNRDLPPYETFEDLREKLLMAVENAQGFEGVD (SEQ ID NO: 9)

WW1(198 – 224):

GWEEKVDNLGR¹⁹⁸TYVNHNNR²²⁴TTQWHRP (SEQ ID NO: 61).

15

WW2 (368 – 396):

PSGWEERKDAKGR³⁶⁸TYVNHNNR³⁹⁶TTTTWTRP (SEQ ID NO: 62).

WW3 (480 – 510):

20 PPGWEMRIAPNGRPFFIDHNTKTTTTWEDPRL (SEQ ID NO: 63).

WW4 (531 – 561):

PPGWEERIHLDGR⁵³¹TFYIDHNSKITQWEDPRL (SEQ ID NO: 64).

25 Human Smurf1 amino acid sequence (uniprot.org/uniprot/ Q9HCE7). The two underlined WW domains correspond to amino acids 234 – 267 (WW1), and 306 – 339 (WW2).

	MSNPGTRRNG	SSIKIRLTVL	CAKNLAKKDF	FRLPDPFAKI	VVDGSGQCHS	50
	TDTVKNTLDP	KWNQHYDLYV	GKTDSTISV	WNHKKIHKKQ	GAGFLGCVRL	100
30	LSNAISRLKD	TGYQRLDLCK	LNPSDTPAVR	GQIVVSLQTR	DRIGTGGSVV	150
	DCRGLLENEG	TVYEDSGPGR	PLSCFMEEPA	PYTDSTGAAA	GGGNCRFVES	200
	PSQDQRLQAO	RLRNPDVGRS	LQTPQNRPHG	<u>HQSP</u> ELPEGY	<u>EQRTTVQ</u> GQV	250
	<u>YFLHTQTGVS</u>	<u>TWHDPRIPSP</u>	<u>SGTIPGGDAA</u>	<u>FLYEFL</u> LQGH	<u>TSEPRD</u> LSNV	300
	<u>NCDELGPLPF</u>	<u>GWEVRSTVSG</u>	<u>RIYFVDHNNR</u>	<u>TTQFTD</u> PRLH	<u>HIMNHQC</u> QLK	350
35	EPSQPLPLPS	EGSLEDEELP	AQRYERDLVQ	KEKVLRRHLS	LQQPQAGHCR	400
	IEVSREEIFE	ESYRQIMKMR	PKDLKKRLMV	KFRGEEGLDY	GGVAREWLYL	450
	LCHEMLNPYY	GLFQYSTONI	YMLQINPDSS	INPDHLSYFH	FVGRIMGLAV	500
	FHGHYINGGF	TVPFYKQLLG	KPIQLSDLES	VDPELHKSLV	WILENDITPV	550
	LDHTFCVEHN	AFGRILQHLEL	KPNGRNVPT	EENKKEYVRL	YVNWRFMGI	600

EAQFLALQKG FNELIPQHLL KPFDQKELEL IIGGLDKIDL NDWKSNTRLK 650
 HCVADSNIWR WFWQAVETFD EERRARLLQF VTGSTRVPLQ GFKALQGSGT 700
 AAGPRLFTIH LIDANTDNLP KAHTCFNRID IPPYESYEKL YEKLLTAVEE 750
 TCGFAVE (SEQ ID NO: 10) 757

5

WW1 (234-267):

PELPEGYEQRRTTVQGQVYFLHTQTGVSTWHDPR (SEQ ID NO: 48).

WW2 (306-339):

10 GPLPPGWEVRSTVSGRIYFVDHNNRTTQFTDPRL (SEQ ID NO: 49).

Human Smurf2 amino acid sequence (uniprot.org/uniprot/Q9HAU4). The three underlined WW domains correspond to amino acids 157 – 190 (WW1), 251 – 284 (WW2), and 297 – 330 (WW3).

15 MSNPGGRRNG FVKRLRLTVLC AKNLVKKDFF RLPDPFAKVV VDCSGQCHST 50
 DTVKNTLDPK WNQHYDLYIG KSDSVTISVW NHKKIHKKQG AGFLGCVRL 100
 SNAINRLKDT GYQRDLCKL GPNDNDTVRG QIVVSLQSRD RGTGGQVVD 150
 CSRLFDNDLP DGWEERRTAS GRIQYLNHIT RTTQWERPTR PASEYSSPGR 200
 PLSCFVDENT PISGTNGATC QSSDPRLAE RRVSQRHRN YMSRTHLHTP 250
 20 PDLPEGYEQR TTQGGQVYFL HTQTGVSTWH DPRVPRDLSN INCEELGPLP 300
 PGWEIRNTAT GRVYFVDHNN RTTQFTDPRL SANLHLVLNR QNQLKDQQQQ 350
 QVVSCLPDDT ECLTVPRYKR DLVQKLKILR QELSQQQPQA GHCRIEVSRE 400
 EIFEESYRQV MKMRPKDLWK RLMIRFRGEE GLDYGGVARE WLYLLSHEML 450
 NPYYGLFQYS RDDIYTLQIN PDSAVNPEHL SYFHFVGRIM GMAVFHGHYI 500
 25 DGGFTLPFYK QLLGKSITLD DMELVDPDLH NSLVWILEND ITGVLDTFC 550
 VEHNAYGEII QHELKPNKGS IPVNEENKKE YVRLYVNWRF LRGIEAQFLA 600
 LQKGFNEVIP QHLLKTFDEK ELELIICGLG KIDVNDWQVN TRLKHCTPDS 650
 NIVKWFWKAV EFFDEERRAR LLQFVTGSSR VPLQGFKALQ GAAGPRLFTI 700
 HQIDACTNNL PKAHTCFNRI DIPYESYEK LYEKLLTAIE ETCGFAVE (SEQ ID NO: 11) 748
 30

WW1 (157-190):

NDLPDGWEERRTASGRIQYLNHITRTTQWERPTR (SEQ ID NO: 50).

WW2 (251-284):

35 PDLPEGYEQRRTTQGGQVYFLHTQTGVSTWHDPRV (SEQ ID NO: 51).

WW3 (297-330):

GPLPPGWEIRNTATGRVYFVDHNNRTTQFTDPRL (SEQ ID NO: 52).

40 Human ITCH amino acid sequence (uniprot.org/uniprot/Q96J02). The four underlined WW domains correspond to amino acids 326 – 359 (WW1), 358 – 391 (WW2), 438 – 471 (WW3), and 478 – 511 (WW4).

MSDSGSQLGS MGSMTMKSQ LITVISAKLK ENKKNWFGPS PYVEVTVDGQ 50
 SKKTEKCNNT NSPKWKQPLT VIVTPVSKLH FRVWSHQTLLK SDVLLGTAAL 100
 45 DIYETLKSNM MKLEEVVVTI QLGGDKPEPE TIGDLSICLD GLQLESEVVT 150
 NGETTCSENG VSLCLPRLEC NSAI SAHCNL CLPGLSDSPI SASRVAGFTG 200
 ASQNDGDSRS KDET RVSTNG SDDPEDAGAG ENRRVSGNNS PLSLNGGFKP 250
 SRPPRPSPRP PPTPRRPASV NGSPSATSES DGSSTGSLPP TNNTNTNSEG 300
 ATSGLIIPLT ISGGSGPRPL NPVTQAPLPP GWEQRVDQHG RYVYVDHVEK 350
 50 RTTWDRPEPL PPGWERRVDN MGRIYYVDHF TRTTTQWRPT LESVRNYEQW 400
 QLQRSQQLGA MQQFNQRFY GNQDLFATSQ SKEFDPLGPL PPGWEKRTDS 450

	NGRVYFVNHN	TRITQWEDPR	SQGQLNEKPL	PEGWEMRFTV	DGIPYFVDHN	500
	RRITTYIDPR	TGKSALDNGP	QIAYVRDFKA	KVQYFRFWCQ	QLAMPQHIKI	550
	TVTRKTLFED	SFQQIMSFSP	QDLRRRLWVI	FPGEGLDYG	GVAREWFFLL	600
	SHEVLNPMYC	LFEYAGKDNY	CLQINPASYI	NPDHLKYFRF	IGRFIAMALF	650
5	HGKFIDTGF	LPFYKRIILNK	PVGLKDLESI	DPEFYNSLIW	VKENNIEECD	700
	LEMYFSVDKE	ILGEIKSHDL	KPNGGNILVT	EENKEEYIRM	VAEWRLSRGV	750
	EEQTQAFFEG	FNEILPQQYL	QYFDAKELEV	LLCGMQEIDL	NDWQRHAIYR	800
	HYARTSKQIM	WFWQFVKEID	NEKRMRLQF	VTGTCRLPVG	GFADLMGSNG	850
	PQKFCIEKVG	KENWLPRSHT	CFNRDLPPY	KSYEQLKEKL	LFAIEETEGF	900
10	GQE	(SEQ ID NO: 12)				903

ITCH WW1 (326-359):

APLPPGWEQRVDQHGRVYYVDHVEKRTTWDRPEP (SEQ ID NO: 53).

15 *ITCH WW2 (358-391):*

EPLPPGWERRVDNMGRIYYVDHFTRTTTWQRPTL (SEQ ID NO: 54).

ITCH WW3 (438-471):

GPLPPGWEKRTDSNGRVYFVNHNTTRITQWEDPRS (SEQ ID NO: 55).

20

ITCH WW4 (478-511):

KPLPEGWEMRFTVDGIPYFVDHNRRTTTYIDPRT (SEQ ID NO: 56).

Human NEDL1 amino acid sequence (uniprot.org/uniprot/Q76N89). The two

25 underlined WW domains correspond to amino acids 829 – 862 (WW1), and 1018 – 1051 (WW2).

	MLLHLCSVK	LYQNRFLGLA	AMASPSRNSQ	SRRRCKEPLR	YSYNPDQFHN	50
	MDLRGGPHDG	VTIPRSTSDT	DLVTSDSRST	LMVSSSYYSI	GHSQDLVIHW	100
	DIKEEVDAGD	WIGMYLIDEV	LSENFLDYKN	RGVNGSHRGQ	IIWKIDASSY	150
30	FVEPETKICF	KYYHGVSGAL	RATTPSVTVK	NSAAPIFKSI	GADETVQGGQ	200
	SRRLISFSL	DFQAMGLRKG	MFFNPDPLYK	ISIQPGKHSI	FPALPHHGQE	250
	RRSKIIGNTV	NPIWQAEQFS	FVSLPTDVLE	IEVKDKFAKS	RPIIKRFLGK	300
	LSMPVQRLLE	RHAIGDRVVS	YTLGRRLLPTD	HVSGQLQFRF	EITSSIHPDD	350
	EEISLSTEPE	SAQIQDSPMN	NLMESGSGEP	RSEAPESSES	WKPEQLGEGS	400
35	VPDGPNGQSI	ELSRPAEEAA	VITEAGDQGM	VSVGPEGAGE	LLAQVQKDIQ	450
	PAPSAEELAE	QLDLGEEASA	LLLEDGEAPA	STKEEPLLEE	ATTQSRAGRE	500
	EEEKEQEEEG	DVSTLEQEG	RLQLRASVKR	KSRPCSLPVS	ELETVIASAC	550
	GDPETPRTHY	IRIHTLLHSM	PSAQGGSAAE	EEDGAEEST	LKDSSEKDGL	600
	SEVDTVAADP	SALEEDREEP	EGATPGTAHP	GHSGGHFPFL	ANGAAQDGD	650
40	HPSTGSESDS	SPRQGGDHSC	EGCDASCCSP	SCYSSSCYST	SCYSSSCYSA	700
	SCYSPSCYNG	NRFASHTRES	SVDSAKISES	TVFSSQDDEE	EENSAFESVP	750
	DSMQSPELDP	ESTNGAGPWQ	DELAAPSGHV	ERSPEGLESP	VAGPSNRREG	800
	ECFILHNSQP	VSQPLSLRPE	HHHYPTIDEP	LPPNWEARID	SHGRVIFYVDH	850
	VNRTTTWQRP	TAAATPDGMR	RSGSIQQMEQ	LNRRYQNIQR	TIATERSEED	900
45	SGSQSCEQAP	AGGGGGGGSD	SEAESSQSSL	DLRREGSLSP	VNSQKITLLL	950
	QSPAVKFITN	PEFFTVLHAN	YSAYRVFTSS	TCLKHMILKV	RRDARNFERY	1000
	QHNRDLVNFI	NMFADTRLEL	PRGWEIKTDQ	QGKSEFFVDHN	SRATTFIDPR	1050
	IPLQNGRLPN	HLTHRQHLQR	LRSYSAGEAS	EVSRNRGASL	LARFGHSLVA	1100
	AIRSQHGHES	LPLAYNDKIV	AFLRQPNIFE	MLQERQPSLA	RNHTLREKIH	1150
50	YIRTEGNHGL	EKLSCDADLV	ILLSLFEEET	MSYVPLQAAF	HPGYSFSPRC	1200
	SPCSSPQNSP	GLQRASARAP	SPYRRDFEAK	LRNFYRKLEA	KGFQGQPGKI	1250
	KLIIRRDHLL	EGTFNQVMAY	SRKELQRNKL	YVTFVGEEL	DYSGPSREFF	1300
	FLLSQELFNP	YYGLFEYSAN	DTYTVQISPM	SAFVENHLEW	FRFSGRILGL	1350
	ALIHQYLLDA	FFTRPFYKAL	LRLPCDLSDL	EYLDDEEFHQ	LQWMKDNNIT	1400
55	DIEDLTFTVN	EEVFGQVTER	ELKSGGANTQ	VTEKNKKEYI	ERMVKWRVER	1450

GVVQQTEALV RGFYEVVDSR LVSVDAREL ELVIAGTAEI DLNDWRNNTTE 1500
 YRGGYHDGHL VIRWFWAAVE RFNNEQRLRL LQFVTGTSSV PYEGFAALRG 1550
 SNGLRRFCIE KWGKITSLEP AHTCFNRLDL PPYPYSMLY EKLLTAVEET 1600
 STFGLE (SEQ ID NO: 13) 1606

5

WW1 (829-862):

PLPPNWEARIDSHGRVFYVDHVNRTTTWQRPTA (SEQ ID NO: 57).

WW2 (1018-1051):

10 LELPRGWEIKTDQQGKSFFVDHNSRATTFIDPRI (SEQ ID NO: 58).

Human NEDL2 amino acid sequence (uniprot.org/uniprot/ Q9P2P5). The two underlined WW domains correspond to amino acids 807 – 840 (WW1), and 985 – 1018 (WW2).

15 MASSAREHLL FVRRRNPMR YTLSPENLQS LAAQSSMPEN MTLQRANSDDT 50
 DLVTSESRSS LTASMYEYTL GQAQNLIIIFW DIKEEVDFSD WIGLYHIDEN 100
 SPANFWD SKN RGVGTGTQKGQ IVWRIEPGPY FMEPEIKICF KYHGISGAL 150
 RATTPCITVK NPAVMMGAEG MEGGASGNLH SRKLVSTLS DLRAVGLKKG 200
 MFFNEDPYLK MSIQPGKKS FPTCAHHCQE RRSTIISNTT NPIWHREKYS 250
 20 FFALLTDVLE IEIKDKFAKS RPIIKRFLGK LTIPVQRLLE RQAIGDQMLS 300
 YNLGRRLPAD HVSGYLQFKV EVTSSVHEDA SPEAVGTILG VNSVNGDLGS 350
 PSDDEDMPGS HHDSQVCSNG PVSEDSAADG TPKHSFRTSS TLEIDTEELT 400
 STSSRTSPFR GRQDSLNDYL DAIEHNGHSR PGTATCSERS MGASPKLRSS 450
 FPTDTRLNAM LHIDSDEEDH EFQQDLGYPS SLEEEGGLIM FSRASRADDG 500
 25 SLTSQTKLED NPVENEEST HEAASFEDKP ENLPDLAESS LPAGFAPEEG 550
 EGGPEPQPSA DQGSALCGS QEVDPQTSQA DTGTSDASGG SRRAVSETES 600
 LDQGSSEPSQV SSETEPSPDA RTESVSEAST RPEGESDLEC ADSSCNESVT 650
 TQLSSVDTRC SSLESARFPE TPAFSSQEEE DGACAAEPTS SGPAEGSQES 700
 VCTAGSLPVV QVPSGEDEGF GAESATVPDQ EELGEVWQRR GSLEGAAAAA 750
 30 ESPPQEEGSA GEAQGTCEGA TAQEEGATGG SQANGHQPLR SLPSVRQDVS 800
 RYQVRDEALP PNWEARIDSH GRIFYVDHVN RITTWQRPTA PPAPQVLQRS 850
 NSIQQMEQLN RRYQSIRRTM TNERPEENTN AIDGAGEEAD FHQASADFRR 900
 ENILPHSTSR SRITLLQSP PVKFLISPEF FTVLHSNPSA YRMFTNNTCL 950
 KHMITKVRD THHFERYQHN RDLVGFLNMF ANKQLELPRG WEMKHDHOGK 1000
 35 AFVVDHNSRT TFIDPRLPL QSSRPTSALV HRQHLTRQRS HSAGEVGEDS 1050
 RHAGPPVLPF PSSTFNTVSR PQYQDMVPVA YNDKIVAFIR QPNIFEILQE 1100
 RQPDLTRNHS LREKIQFIRT EGTPLGLVRLS SDADLVMLLS LFEEEIMSYV 1150
 PPHALLHPSY CQSPRGSPVS SPQNSPGTQR ANARAPAPYK RDEAKLRNF 1200
 YRKLETGKYG QGPGKLLII RRDHLEDAF NQIMGYSRKD LQRNKLYVTF 1250
 40 VGEEGLDYSY PSREFFFLVS RELFNPYYGL FEYSANDTYT VQISPMASAFV 1300
 DNHHEWFRFS GRILGLALIH QYLLDAFFTR PFYKALLRIL CDLSLEYLD 1350
 EEHFQSLQWM KDNDIHDILD LTFTVNNEEVF GQITERELKP GGANIPVTEK 1400
 NKKEYIERMV KWRIERGVSQ QTESLVRCFY EVVDARLVSV FDARELELVI 1450
 AGTAEIDLSD WRNNTYRGG YHDNHIVIRW FWAAVERFNN EQRLRLQLFV 1500
 45 TGTSSIPYEG FASLRGSGNP RRFCVEKWGK ITALPRAHTC FNRLDLPPYP 1550
 SFSMLYEKLL TAVEETSTFG LE (SEQ ID NO: 14) 1572

WW1 (807-840):

EALPPNWEARIDSHGRIFYVDHVNRTTTWQRPTA (SEQ ID NO: 59).

50

WW2 (985-1018):

LELPRGWEMKHDHOGKAFFVDHNSRTTTFIDPRL (SEQ ID NO: 60).

In some embodiments, the WW domain comprises a WW domain or WW domain variant from the amino acid sequence (SEQ ID NO: 6); (SEQ ID NO: 7); (SEQ ID NO: 8); (SEQ ID NO: 9); (SEQ ID NO: 10); (SEQ ID NO: 11); (SEQ ID NO: 12); (SEQ ID NO: 13); or (SEQ ID NO: 14). In other embodiments, the WW domain consists of a WW domain or WW domain variant from the amino acid sequence (SEQ ID NO: 6); (SEQ ID NO: 7); (SEQ ID NO: 8); (SEQ ID NO: 9); (SEQ ID NO: 10); (SEQ ID NO: 11); (SEQ ID NO: 12); (SEQ ID NO: 13); or (SEQ ID NO: 14). In another embodiment, the WW domain consists essentially of a WW domain or WW domain variant from the amino acid sequence (SEQ ID NO: 6); (SEQ ID NO: 7); (SEQ ID NO: 8); (SEQ ID NO: 9); (SEQ ID NO: 10); (SEQ ID NO: 11); (SEQ ID NO: 12); (SEQ ID NO: 13); or (SEQ ID NO: 14). Consists essentially of means that a domain, peptide or polypeptide consists essentially of an amino acid sequence when such an amino acid sequence is present with only a few additional amino acid residues, for example, from about 1 to about 10 or so additional residues, typically from 1 to about 5 additional residues in the domain, peptide or polypeptide.

Alternatively, the WW domain may be a WW domain that has been modified to include two basic amino acids at the C-terminus of the domain. Techniques are known in the art and are described in the art, for example, in Sambrook *et al.* ((2001) *Molecular Cloning: a Laboratory Manual*, 3rd ed., Cold Spring Harbour Laboratory Press). Thus, a skilled person could readily modify an existing WW domain that does not normally have two C-terminal basic residues so as to include two basic residues at the C-terminus.

Basic amino acids are amino acids that possess a side-chain functional group that has a pKa of greater than 7 and include lysine, arginine, and histidine, as well as basic amino acids that are not included in the twenty α -amino acids commonly included in proteins. The two basic amino acids at the C-terminus of the WW domain may be the same basic amino acid or may be different basic amino acids. In one embodiment, the two basic amino acids are two arginines.

The term WW domain also includes variants of a WW domain provided that any such variant possesses two basic amino acids at its C-terminus and maintains the ability of the WW domain to associate with the PPXY motif. A variant of such a WW domain refers to a WW domain which retains the ability to associate with the PPXY motif (*i.e.*, the PPXY motif of minimal ARRDC1) and that has been mutated at one or more amino acids, including point, insertion or deletion mutations, but still retains the ability to associate with the PPXY motif. A variant or derivative therefore includes deletions, including truncations and fragments; insertions and additions, for example, conservative substitutions, site-directed mutants and

allelic variants; and modifications, including one or more non-amino acyl groups (*e.g.*, sugar, lipid, etc.) covalently linked to the peptide and post-translational modifications. In making such changes, substitutions of like amino acid residues can be made on the basis of relative similarity of side-chain substituents, for example, their size, charge, hydrophobicity, hydrophilicity, and the like, and such substitutions may be assayed for their effect on the function of the peptide by routine testing.

The WW domain may be part of a longer protein. Thus, the protein, in various different embodiments, comprises the WW domain, consists of the WW domain or consists essentially of the WW domain, as defined herein. The polypeptide may be a protein that includes a WW domain as a functional domain within the protein sequence. In some embodiments, the polypeptide is an endonuclease. In some embodiments, the endonuclease is Cas9 protein or a Cas9 protein variant. In other embodiments, the polypeptide comprises the sequence set forth in (SEQ ID NO: 6); (SEQ ID NO: 7); (SEQ ID NO: 8); (SEQ ID NO: 9); (SEQ ID NO: 10); (SEQ ID NO: 11); (SEQ ID NO: 12); (SEQ ID NO: 13); or (SEQ ID NO: 14), consists of (SEQ ID NO: 6); (SEQ ID NO: 7); (SEQ ID NO: 8); (SEQ ID NO: 9); (SEQ ID NO: 10); (SEQ ID NO: 11); (SEQ ID NO: 12); (SEQ ID NO: 13); or (SEQ ID NO: 14), or consists essentially of (SEQ ID NO: 6); (SEQ ID NO: 7); (SEQ ID NO: 8); (SEQ ID NO: 9); (SEQ ID NO: 10); (SEQ ID NO: 11); (SEQ ID NO: 12); (SEQ ID NO: 13); or (SEQ ID NO: 14).

The term "target site," as used herein in the context of functional effector proteins that bind a nucleic acid molecule, such as nucleases, deaminases, and transcriptional activators or repressors, refers to a sequence within a nucleic acid molecule that is bound and acted upon by the effector protein, *e.g.*, cleaved by the nuclease or transcriptionally activated or repressed by the transcriptional activator or repressor, respectively. A target site may be single-stranded or double-stranded. In the context of RNA-guided (*e.g.*, RNA-programmable) nucleases (*e.g.*, a protein dimer comprising a Cas9 gRNA binding domain and an active Cas9 DNA cleavage domain), a target site typically comprises a nucleotide sequence that is complementary to the gRNA of the RNA-programmable nuclease, and a protospacer adjacent motif (PAM) at the 3' end adjacent to the gRNA-complementary sequence. For the RNA-guided nuclease Cas9, the target site may be, in some embodiments, 20 base pairs plus a 3 base pair PAM (*e.g.*, NNN, wherein N represents any nucleotide). Typically, the first nucleotide of a PAM can be any nucleotide, while the two downstream nucleotides are specified depending on the specific RNA-guided nuclease. Exemplary target sites for RNA-guided nucleases, such as Cas9, are known to those of skill in the art and

include, without limitation, NNG, NGN, NAG, and NGG, wherein N represents any nucleotide. In addition, Cas9 nucleases from different species (*e.g.*, *S. thermophilus* instead of *S. pyogenes*) recognizes a PAM that comprises the sequence NGGNG. Additional PAM sequences are known, including, but not limited to, NNAGAAW (SEQ ID NO: 121) and

5 NAAR (see, *e.g.*, Esvelt and Wang, *Molecular Systems Biology*, 9:641 (2013), the entire contents of which are incorporated herein by reference). For example, the target site of an RNA-guided nuclease, such as, *e.g.*, Cas9, may comprise the structure [NZ]-[PAM], where each N is, independently, any nucleotide, and Z is an integer between 1 and 50, inclusive. In some embodiments, Z is at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least

10 8, at least 9, at least 10, at least 11, at least 12, at least 13, at least 14, at least 15, at least 16, at least 17, at least 18, at least 19, at least 20, at least 25, at least 30, at least 35, at least 40, at least 45, or at least 50. In some embodiments, Z is 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or 50. In some embodiments, Z is 20. In some embodiments,

15 “target site” may also refer to a sequence within a nucleic acid molecule that is bound but not cleaved by a nuclease. For example, certain embodiments described herein provide proteins comprising an inactive (or inactivated) Cas9 DNA cleavage domain. Such proteins (*e.g.*, when also including a Cas9 RNA binding domain) are able to bind the target site specified by the gRNA, however because the DNA cleavage site is inactivated, the target site is not

20 cleaved by the particular protein. However, such proteins as described herein are typically associated with another protein (*e.g.*, a nuclease or transcription factor) or molecule that mediates cleavage of the nucleic acid molecule. In some embodiments, the sequence actually cleaved will depend on the protein (*e.g.*, nuclease) or molecule that mediates cleavage of the nucleic acid molecule, and in some cases, for example, will relate to the proximity or distance

25 from which the inactivated Cas9 protein(s) is/are bound.

Detailed Description of Certain Embodiments of the Invention

The instant disclosure relates to the discovery that a minimal ARRDC1 protein can efficiently package cargos. Overexpression of a minimal ARRDC1 can produce larger

30 amounts of functional ARRDC1 than overexpression of full length ARRDC1, while still achieving packaging of cargos into ARMMs. Minimal ARRDC1 constructs may reduce the volume of ARRDC1 in ARMMs, thus increasing the loading capacity of an ARMM. Reducing the size of the ARRDC1 required to achieve packaging of cargo also increases the practical limit of cargos that can be expressed with minimal ARRDC1 as a direct fusion or

linked to the minimal ARRDC1 molecule post-translationally. Motifs in the minimal ARRDC1 protein yield efficient ARMM budding: the arrestin domain directs the protein to the plasma membrane; the tetrapeptide PSAP (SEQ ID NO: 122) or PTAP (SEQ ID NO: 123) interacts with and recruits TSG101 and ESCRT I complex proteins to the plasma membrane; and the PPXY motif(s) interact with the WW domains of NEDD4 E3 ligases to enhance ARMM budding (Nabhan et al., 2012).

Aspects of the present disclosure provide a minimal ARRDC1 that is shorter than the full-length ARRDC1 protein, yet maintains its same function with respect to microvesicle formation. In some embodiments, the minimal ARRDC1 comprises at least a portion of an arrestin domain, at least one PSAP (SEQ ID NO: 122) or PTAP (SEQ ID NO: 123) motif, and at least one PPXY motif, wherein the minimal ARRDC1 is shorter than the full-length ARRDC1 protein. A non-limiting example of a minimal ARRDC1 is provided in SEQ ID NO: 1.

MGRVQLFEISLSHGRVVYSPGEPLAGTVRVRLGAPLPFRAIRVTCIGSCGVSNKANDT
AWVVEEGYFNSSLADKGSLPAGEHSFPFQFLLPATAPTSFEGPFGKIVHQVRAAIHT
PRFSKDHKCSLVFYILSPLNLNSIPDIEQPNVASATKKFSYKLVKTGSSVLTASTDLRGY
VVGQALQLHADVENQSGKDTSPVVASLLQKVSYKAKRWIHDVRTIAEVEGAGVKA
WRRAQWHEQILVPALPQSALPGCSLIHIDYYLQVSLKAPEATVTLPVFIGNIAVNHAP
VSPRPGLGLPPGAPPLVVPSAPPQEEAEPPEYPYEAPPSY (SEQ ID NO: 1). Another

non-limiting example of a minimal ARRDC1 is provided in SEQ ID NO: 125:

MGRVQLFEISLSHGRVVYSPGEPLAGTVRVRLGAPLPFRAIRVTCIGSCGVSNKANDT
AWVVEEGYFNSSLADKGSLPAGEHSFPFQFLLPATAPTSFEGPFGKIVHQVRAAIHT
PRFSKDHKCSLVFYILSPLNLNSIPDIEQPNVASATKKFSYKLVKTGSSVLTASTDLRGY
VVGQALQLHADVENQSGKDTSPVVASLLQKVSYKAKRWIHDVRTIAEVEGAGVKA
WRRAQWHEQILVPALPQSALPGCSLIHIDYYLQVSLKAPEATVTLPVFIGNIAVNHAP
VSPRPGLGLPPGAPPLVVPTAPPQEEAEPPEYPYEAPPSY (SEQ ID NO: 125).

A non-limiting example of a nucleic acid sequence encoding a minimal ARRDC1 is provided in SEQ ID NO: 126:

Atggggcgagtgagctcttcgagatcagcctgagccacggccgctcgtctacagccccggggagccgttggtgggaccgtgc
gcgtgcgcctgggggacccgtgccgtccgagccatccgggtgacctgcataggtcctgcggggtctccaacaaggctaatacaca
cagcgtgggttagtgaggagggttacttcaacagttccctgtcgtgagcagacaaggggagcctgcccgttgagagcacagcttc
ccctccagttcctgcttctgccactgcacccacgtcctttgagggtccttcgggaagatcgtgcaccaggtgaggggcccatcca

cacgccacgggtttccaaggatcacaaagtgcagcctcgtgttctatattctgagcccttgaacctgaacagcatcccagacattgagca
 acccaacgtggcctctgccaccaagaagtctctacaagctggtgaagacgggcagcgtggctctcacagccagcactgatctccg
 cggctatgtggtggggcaggcactgcagctgcatgccgacgttgagaaccagtcaggcaaggacaccagccctgtggtggccagtc
 tgctgcagaaagtgtcctataaggccaagcgtggatccacgacgtacggaccattgcggaggtggaggggtcggggcgtcaaggc
 5 ctggcgggcgggcgagtgccacgagcagatcctggtgcctgccttgcgccagtcggccctgccgggctgcagcctcatccacatcg
 actactacttacaggtctctctgaaggcgccggaagctactgtgacctcccggtcttcattggcaatattgctgtgaacctgccccagt
 gagccccggccaggcctggggctgcctcctggggccccaccctggtggtgccttccgcaccaccccaggaggaggtgagcc
 ctatgaggccccaccgtcttat (SEQ ID NO: 126)

A minimal ARRDC1 may comprise an amino acid sequence having at least 80%, at
 10 least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at
 least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at
 least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identity to SEQ ID NO: 1.
 In some embodiments, the minimal ARRDC1 comprises SEQ ID NO: 1. In some
 embodiments, the minimal ARRDC1 consists of SEQ ID NO: 1.

15 A minimal ARRDC1, as provided herein, may comprise any length amino acid
 sequence that is fewer amino acids than full-length ARRDC1. In some embodiments, the
 minimal ARRDC1 comprises between 100-400 amino acids. In some embodiments, the
 minimal ARRDC1 comprises between 100-350 amino acids, between 100-300 amino acids,
 between 100-250 amino acids, between 100-200 amino acids, or between 100-150 amino
 20 acids. In some embodiments, the minimal ARRDC1 comprises between 150-350 amino
 acids, between 200-350 amino acids, between 250-350 amino acids, or between 300-350
 amino acids.

In some embodiments, the minimal ARRDC1 comprises up to 400 amino acids, up
 375 amino acids, up to 350 amino acids, up to 325 amino acids, up to 300 amino acids, up to
 25 275 amino acids, up to 250 amino acids, up to 225 amino acids, up to 200 amino acids, up to
 200 amino acids, up to 175 amino acids, up to 150 amino acids, up to 125 amino acids, or up
 to 100 amino acids.

In some embodiments, the minimal ARRDC1 comprises about 400 amino acids,
 about 375 amino acids, about 350 amino acids, about 325 amino acids, about 300 amino
 30 acids, about 275 amino acids, about 250 amino acids, about 225 amino acids, about 200
 amino acids, about 175 amino acids, about 150 amino acids, about 125 amino acids, or about
 100 amino acids.

In some embodiments, the minimal ARRDC1 comprises at least a portion of an arrestin domain. In some embodiments, the portion of the arrestin domain comprises amino acids 1-308 of SEQ ID NO: 116.

Minimal ARRDC1 as provided herein may comprise any number of PSAP (SEQ ID NO: 122) motifs and/or PTAP (SEQ ID NO: 123) motifs. In some embodiments, the minimal ARRDC1 comprises at least one PSAP (SEQ ID NO: 122) motif or at least one PTAP (SEQ ID NO: 123) motif. In some embodiments, the minimal ARRDC1 comprises at least one PSAP (SEQ ID NO: 122) motif. In some embodiments, the minimal ARRDC1 comprises at least one PTAP (SEQ ID NO: 123) motif. In some embodiments, the minimal ARRDC1 comprises at least one PSAP (SEQ ID NO: 122) motif and at least one PTAP (SEQ ID NO: 123) motif.

Minimal ARRDC1 as provided herein may comprise any number of PPXY motifs. In some embodiments, the minimal ARRDC1 comprises at least one PPXY motif. In some embodiments, the minimal ARRDC1 comprises at least two PPXY motifs. In some embodiments, the minimal ARRDC1 comprises at least three PPXY motifs.

The minimal ARRDC1, in some embodiments, encompasses functional variants. A functional variant may contain one or more mutations outside the functional domain(s) of the minimal ARRDC1, for example, a mutation outside the arrestin domain, or a mutation within the arrestin domain which does not affect its function. Mutations outside the functional domain(s) would not be expected to affect the biological activity of the protein. For example, mutation outside the functional domain would not be expected to substantially affect the formation or budding of microvesicles. In some embodiments, a functional variant may comprise an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identity to SEQ ID NO: 1.

Alternatively or in addition, the functional mutation may contain a conservative mutation(s) at one or more positions in the minimal ARRDC1. For example, the functional variant may contain a conservative mutation at up to 20 positions, up to 15 positions, up to 10 positions, up to 5 positions, up to 4 positions, up to 3 positions, up to 2 positions, or only at 1 position.

Microvesicles with WW-domain-containing-cargos

Some aspects of this invention provide arrestin domain-containing protein 1 (ARRDC1)-mediated microvesicles (ARMMs) containing a cargo that is fused to a WW domain. Such ARMMs typically include a lipid bilayer and a minimal ARRDC1 protein or variant thereof. In some embodiments, the cargo is fused to a WW domain that associates with the PPXY (where x is any amino acid) domain of minimal ARRDC1 which may facilitate loading of the cargo into an ARMM. In some embodiments, the cargo is a protein, nucleic acid, or small molecule. In some embodiments, the cargo is a Cas9 protein or Cas9 variant. In some embodiments the Cas9 protein or variant is a fusion protein. For example, the Cas9 protein or Cas9 variant may be fused to one or more WW domains to facilitate loading into an ARMM. In some embodiments, the Cas9 fusion protein or Cas9 variant is fused to one or more nuclear localization sequences (NLSs) to facilitate translocation of the Cas9 fusion protein into the nucleus of a target cell. In certain embodiments the Cas9 variant is a Cas9 protein or Cas9 protein variant comprising an active or inactive DNA cleavage domain of Cas9 or a partially inactive DNA cleavage domain (*e.g.*, a Cas9 “nickase”), and/or the gRNA binding domain of Cas9. It should be appreciated that any number of proteins, nucleic acids, or small molecules known in the art can be fused to one or more WW domains to generate a cargo that can be loaded into an ARMM, for example, a reprogramming factor (*e.g.*, Oct4, Sox2, c-Myc, or KLF4) may be fused to one or more WW domains to facilitate loading of one or more reprogramming factors into an ARMM. In some embodiments, the cargo protein is a therapeutic protein (*e.g.*, a transcription factor, a tumor suppressor, a developmental regulator, a growth factor, a metastasis suppressor, a pro-apoptotic protein, a zinc finger nuclease, or a recombinase) that is fused to one or more WW domains. In other embodiments, an ARMM further includes a non-cargo protein, such as a TSG101 protein or variant thereof to facilitate the release of ARMMs. The TSG101 protein interacts with ARRDC1, which results in relocation of TSG101 from endosomes to the plasma membrane and mediates the release of microvesicles that contain TSG101, ARRDC1, and other cellular components, including, for example, cargo proteins, nucleic acids (*i.e.*, gRNAs), and small molecules.

In some embodiments, microvesicle, *e.g.*, ARMMs, are provided that comprise a minimal ARRDC1 protein fragment, and/or a TSG101 protein fragment. In some embodiments, fusion proteins are provided that comprise a minimal ARRDC1 protein fragment and/or a TSG101 protein fragment. In some embodiments, expression constructs are provided that encode a minimal ARRDC1 protein fragment and/or a TSG101 protein

fragment. In some embodiments, the minimal ARRDC1 protein fragment is a C-terminal minimal ARRDC1 protein fragment. In some embodiments, the ARRDC1 protein fragment comprises the PSAP (SEQ ID NO: 122) motif and at least 10, at least 20, at least 30, at least 40, at least 50, at least 60, at least 70, at least 80, at least 90, at least 100, at least 110, at least 120, at least 130, at least 140, at least 150, at least 160, at least 170, at least 180, at least 190, at least 200, at least 210, at least 220, at least 230, at least 240, at least 250, at least 260, at least 270, at least 280, at least 290, or at least 300 contiguous amino acids of the minimal ARRDC1 sequence. In some embodiments, the TSG101 protein fragment comprises a TSG101 UEV domain. In some embodiments, the TSG101 protein fragment comprises the UEV domain and comprises at least 150, at least 160, at least 170, at least 180, at least 190, at least 200, at least 210, at least 220, at least 230, at least 240, at least 250, at least 260, at least 270, at least 280, at least 290, or at least 300 contiguous amino acids of the TSG101 sequence.

In some embodiments, the inventive microvesicles, e.g., ARMMs comprising a minimal ARRDC1, further comprise a cell surface protein, for example, an integrin, a receptor tyrosine kinase, a G-protein coupled receptor, or a membrane-bound immunoglobulin. Other cell surface proteins may also be included in an ARMM. Integrins, receptor tyrosine kinases, G-protein coupled receptors, and a membrane-bound immunoglobulins suitable for use with embodiments of this invention will be apparent to those of skill in the art and the invention is not limited in this respect. For example, in some embodiments, the integrin is an $\alpha 1\beta 1$, $\alpha 2\beta 1$, $\alpha 4\beta 1$, $\alpha 5\beta 1$, $\alpha 6\beta 1$, $\alpha L\beta 2$, $\alpha M\beta 2$, $\alpha IIb\beta 3$, $\alpha V\beta 3$, $\alpha V\beta 5$, $\alpha V\beta 6$, or $\alpha 6\beta 4$ integrin. In some embodiments, the receptor tyrosine kinase is an EGF receptor (ErbB family), insulin receptor, PDGF receptor, FGF receptor, VEGF receptor, HGF receptor, Trk receptor, Eph receptor, AXL receptor, LTK receptor, TIE receptor, ROR receptor, DDR receptor, RET receptor, KLG receptor, RYK receptor, or MuSK receptor. In some embodiments, the G-protein coupled receptor is a rhodopsin-like receptor, the secretin receptor, metabotropic glutamate/pheromone receptor, cyclic AMP receptor, frizzled/smoothed receptor, CXCR4 receptor, CCR5 receptor, or beta-adrenergic receptor.

Some aspects of this invention relate to the recognition that ARMMs are taken up by target cells, and ARMM uptake results in the release of the contents of the ARMM into the cytoplasm of the target cells. Some aspects of this invention relate to the recognition that this can be used to deliver an agent in ARMMs to the target cell or a population of target cells, for example, by contacting the target cell with ARMMs comprising the agent to be delivered.

Accordingly, some aspects of this invention provide ARMMs that comprise an agent, for example, a recombinant nucleic acid, a recombinant protein, or a synthetic small molecule.

In some embodiments, the agent is an agent that effects a desired change in the target cell, for example, a change in cell survival, proliferation rate, a change in differentiation stage, a change in a cell identity, a change in chromatin state, a change in the transcription rate of one or more genes, a change in the transcriptional profile, or a post-transcriptional change in gene expression of the target cell. It will be understood by those of skill in the art, that the agent to be delivered will be chosen according to the desired effect in the target cell. For example, to effect a change in the differentiation stage of a target cell, for example, to reprogram a differentiated target cell into an embryonic stem cell-like stage, the cell is contacted, in some embodiments, with ARMMs with reprogramming factors, for example, Oct4, Sox2, c-Myc, and/or KLF4. Similarly, to effect the change in the chromatin state of a target cell, the cell is contacted, in some embodiments, with ARMMs containing a chromatin modulator, for example, a DNA methyltransferase or a histone deacetylase. For another example, if survival of the target cell is to be diminished, the target cell, in some embodiments, is contacted with ARMMs comprising a cytotoxic agent, for example, a chemotherapeutic drug. Additional agents suitable for inclusion into ARMMs and for a ARMM-mediated delivery to a target cell or target cell population will be apparent to those skilled in the art, and the invention is not limited in this respect.

In some embodiments, the agent is included in the ARMMs by contacting cells producing the ARMMs with the agent. For example, if the agent is a small molecule, for example a therapeutic drug to be delivered to a target cell population within the body of a subject, ARMMs containing the drug are produced by contacting cells expressing minimal ARRDC1 with the drug in an amount and for a time sufficient to generate ARMMs containing the drug. For another example, if the agent is a nucleic acid or a protein, ARMMs containing nucleic acid or the protein are produced by expressing the nucleic acid or the protein in cells expressing minimal ARRDC1 and TSG101, for example, from a recombinant expression construct.

In some embodiments, the agent is conjugated to the minimal ARRDC1 protein, the minimal ARRDC1 fragment, the TSG101 protein, or the TSG101 fragment. In some embodiments, where the agent is a protein, the protein may be conjugated to the ARRDC protein, the minimal ARRDC1 fragment, the TSG101 protein, or the TSG101 fragment, by expressing the protein agent as a fusion with the ARRDC1 protein, the ARRDC1 fragment, the TSG101 protein, or the TSG101 fragment.

In some embodiments, ARMMs comprising a minimal ARRDC1 are provided that include a recombinant or a synthetic nucleic acid. Such ARMMs can be used to deliver the recombinant or synthetic nucleic acids to a target cell or target cell population. In some embodiments, the recombinant nucleic acid comprises an RNA, for example, an RNA
5 encoding a protein (e.g., an mRNA), or a non-coding RNA. In some embodiments, the nucleic acid comprises an RNAi agent, for example, an antisense RNA, a small interfering RNA (siRNA), a small hairpin RNA (shRNA), a microRNA (miRNA), a small nuclear RNA (snRNA), a small nucleolar RNA (snoRNA), or a long intergenic non-coding RNA (lincRNA), or a precursor thereof. Some embodiments, ARMMs are provided that include a
10 recombinant structural RNA, a ribozyme, or a precursor thereof.

Coding RNAs, RNAi agents, structural RNAs, and ribozymes, as well as precursors thereof, are well known to those skilled in the art and suitable RNAs and RNAi agents according to aspects of this invention will be apparent to the skilled artisan. It will be appreciated that the invention is not limited in this respect. ARMMs including RNA can be
15 used to express the RNA function in a target cell without the need for genetic manipulation of the target cell. For example, ARMMs including protein-encoding nucleic acids can be used to express the encoded protein in a target cell or cell population upon ARMMs uptake without the need to genetically manipulate the target cell or cell population. For another example, ARMMs including an RNAi agent can be used to knock down a gene of interest in
20 the target cell or the target cell population without the need to genetically amended claims department cell or cell population. For a third example, ARMMs including a ribozyme can be used to modulate the expression of a target nucleic acid, or to edit a target mRNA and a target cell without the need for genetic manipulation.

In some embodiments, ARMMs comprising a minimal ARRDC1 are provided that
25 include a DNA, for example, a vector including an expression construct, a LINE sequence, a SINE sequence, a composite SINE sequence, or an LTR-retrotransposon sequence. ARMMs containing DNA allow for the transfer of genes or DNA elements from cell to cell, or, in some embodiments, for the targeted insertion of genes or DNA elements into a target cell or target cell type, for example a pathological target cell type in a subject. In some
30 embodiments, ARMMs are provided that include a DNA encoding a protein. In some embodiments, ARMMs are provided that include a DNA encoding a non-coding RNA, for example, an antisense RNA, a small interfering RNA (siRNA), a small hairpin RNA (shRNA), a microRNA (miRNA), a small nuclear RNA (snRNA), a small nucleolar RNA (snoRNA), or a long intergenic non-coding RNA (lincRNA), or a precursor thereof. In some

embodiments, the use of ARMMs containing a DNA has the advantage that a higher level of expression or a more sustained expression of the encoded protein or RNA can be achieved in a target cell as compared to direct delivery of the protein or RNA. In some embodiments, the DNA included in the ARMMs comprises a cell type specific promoter controlling the
5 conscription of the encoded protein or RNA. The use of a cell type specific promoter allows for the targeted expression of the proteins were RNA encoded by the ARMM-delivered DNA, which can be used, for example in some therapeutic embodiments, to minimize the effect on subpopulations that are not targeted but may take up ARMMs.

In some embodiments, ARMMs comprising a minimal ARRD1 are provided that
10 include a detectable label. Such ARMMs allow for the labeling of a target cell without genetic manipulation. Detectable labels suitable for direct delivery to target cells are known in the art, and include, but are not limited to, fluorescent proteins, fluorescent dyes, membrane-bound dyes, and enzymes, for example, membrane-bound enzymes, catalyzing the reaction resulting in a detectable reaction product. Detectable labels suitable according to
15 some aspects of this invention further include membrane-bound antigens, for example, membrane-bound ligands that can be detected with commonly available antibodies or antigen binding agents.

In some embodiments, ARMMs are provided that comprise a therapeutic agent. It will be appreciated, that any therapeutic agent that can be introduced into a cell shedding
20 ARMMs or that can be packaged into synthetic ARMMs is suitable for inclusion into ARMMs according to some aspects of this invention. Suitable therapeutic agents include, but are not limited to, small organic molecules, also referred to as small molecules, or small compounds, and biologics, for example, therapeutic proteins, or protein fragments. Some non-limiting examples of therapeutic agents suitable for inclusion in ARMMs include
25 antibacterial agents, antifungal antibiotics, antimyobacterials, neuraminidase inhibitors, antineoplastic agents, cytotoxic agents, cholinergic agents, parasympathomimetics, anticholinergic agents, antidepressants, antipsychotics, respiratory and cerebral stimulants, proton pump inhibitors, hormones and synthetic substitutes, receptor ligands, kinase inhibitors, chemotherapeutic agents, signaling molecules, kinases, phosphatases, proteases,
30 RNA editing enzymes, nucleases, and zinc finger proteins.

In some embodiments, ARMMs are provided that comprise a protein to be delivered to a target cell. In some embodiments, the protein is or comprises a transcription factor, a transcriptional repressor, a fluorescent protein, a kinase, a phosphatase, a protease, a ligase, a chromatin modulator, or a recombinase. In some embodiments, the protein is a therapeutic

protein. In some embodiments the protein is a protein that effects a change in the state or identity of a target cell. For example, in some embodiments, the protein is a reprogramming factor. Suitable transcription factors, transcriptional repressors, fluorescent proteins, kinases, phosphatases, proteases, ligases, chromatin modulators, recombinases, and reprogramming
5 factors are known to those skilled in the art, and the invention is not limited in this respect.

In some embodiments, ARMMs are provided that comprise an agent, for example, a small molecule, a nucleic acid, or a protein, that is covalently or non-covalently bound, or conjugated, to a minimal ARRDC1 protein or fragment thereof, or a TSG101 protein or fragment thereof. In some embodiments, agent is conjugated to the minimal ARRDC1
10 protein or fragment thereof, or the TSG101 protein or fragment thereof, via a linker. The linker may be cleavable or uncleavable. In some embodiments, the linker comprises an amide, ester, ether, carbon-carbon, or disulfide bond, although any covalent bond in the chemical art may be used. In some embodiments, the linker comprises a labile bond, cleavage of which results in separation of the supercharged protein from the peptide or
15 protein to be delivered. In some embodiments, the linker is cleaved under conditions found in the target cell (e.g., a specific pH, a reductive environment, or the presence of a cellular enzyme). In some embodiments, the linker is cleaved by an enzyme, for example, a cellular enzyme. In some embodiments, the enzyme is a cellular protease or a cellular esterase. In some embodiments, the cellular protease is a cytoplasmic protease, an endosomal protease, or
20 an endosomal esterase. In some embodiments, the cellular enzyme is specifically expressed in a target cell or cell type, resulting in preferential or specific release of the functional protein or peptide in the target cell or cell type. The target sequence of the protease may be engineered into the linker between the agent to be delivered and the minimal ARRDC1 protein or the TSG101 protein or fragment thereof. In some embodiments, the target cell or
25 cell type is a cancer cell or cancer cell type, a cell or cell type of the immune system, or a pathologic or diseased cell or cell type, and the linker is cleaved by an enzyme or based on a characteristic specific for the target cell. In some embodiments, the linker comprises an amino acid sequence chosen from the group including AGVF (SEQ ID NO: 114), GFLG (SEQ ID NO: 117), FK, AL, ALAL (SEQ ID NO: 118), or ALALA (SEQ ID NO: 119).
30 Other suitable linkers will be apparent to those of skill in the art. In some embodiments, the linker is a cleavable linker. In some embodiments, the linker comprises a protease recognition site. In certain embodiments, the linker is a UV-cleavable moiety. Suitable linkers, for example, linkers comprising a protease recognition site, or linkers comprising a UV cleavable moiety are known to those of skill in the art. In some embodiments, the agent is

conjugated to the minimal ARRDC1 protein or fragment thereof via a sortase reaction, and the linker comprises an LPXTG (SEQ ID NO: 120) motif. Methods and reagents for conjugating agents according to some aspects of this invention to proteins are known to those of skill in the art. Accordingly, suitable method for conjugating and agents to be included in
5 an ARMM to an minimal ARRDC1 protein or fragment thereof, or a TSG101 protein or fragment thereof will be apparent to those of skill in the art based on this disclosure.

Methods for isolating ARMMs are also provided herein. One exemplary method includes collecting the culture medium, or supernatant, of a cell culture comprising microvesicle-producing cells. In some embodiments, the cell culture comprises cells
10 obtained from a subject, for example, cells suspected to exhibit a pathological phenotype, e.g., a hyperproliferative phenotype. In some embodiments, the cell culture comprises genetically engineered cells producing ARMMs, for example, cells expressing a recombinant ARMM protein, for example, a recombinant minimal ARRDC1 or TSG101 protein, such as a minimal ARRDC1 or TSG101 fusion protein. In some embodiments, the supernatant is pre-
15 cleared of cellular debris by centrifugation, for example, by two consecutive centrifugations of increasing G value (e.g., 500G and 2000G). In some embodiments, the method comprises passing the supernatant through a 0.2 μ m filter, eliminating all large pieces of cell debris and whole cells. In some embodiments, the supernatant is subjected to ultracentrifugation, for example, at 120,000g for 2 h, depending on the volume of centrifugate. The pellet obtained
20 comprises microvesicles. In some embodiments, exosomes are depleted from the microvesicle pellet by staining and/or sorting (e.g., by FACS or MACS) using an exosome marker as described herein. Isolated or enriched ARMMs can be suspended in culture media or a suitable buffer, as described herein.

25 *WW domain containing cargos*

Aspects of the disclosure relate to ARMMs comprising a cargo associated with at least one WW domain. In some aspects, fusion proteins are provided that comprise a cargo protein with at least one WW domain. In some aspects, expression constructs are provided that encode a cargo protein associated with at least one WW domain. The WW domain of a
30 cargo protein may associate with the PPXY motif of ARRDC1, or variant thereof, to facilitate association with or inclusion of the cargo protein into an ARMM. A schematic representation of a Cas9 cargo protein fused to a WW domain that associates with the PPXY motif of ARRDC1 can be seen in Figure 2. In some embodiments, the cargo protein is fused to at least one, at least two, at least three, at least four, at least five, at least six, at least seven,

at least eight, at least nine, at least ten, or more WW domains. The WW domain may be derived from a WW domain of the ubiquitin ligase WWP1, WWP2, Nedd4-1, Nedd4-2, Smurf1, Smurf2, ITCH, NEDL1, or NEDL2 (Fig. 1). For example, the WW domain may comprise a WW domain or WW domain variant from the amino acid sequence set forth in

5 (SEQ ID NO: 6); (SEQ ID NO: 7); (SEQ ID NO: 8); (SEQ ID NO: 9); (SEQ ID NO: 10); (SEQ ID NO: 11); (SEQ ID NO: 12); (SEQ ID NO: 13); or (SEQ ID NO: 14). In certain embodiments, the cargo proteins may comprise two WW domains or WW domain variants from the human ITCH protein having the amino acid sequence:

PLPPGWEQRVDQHGRVYYVDHVEKRTTWDRPEPLPPGWERRVDNMGRIYYVDHFT

10 RTTTWQRPTL (SEQ ID NO: 18).

In other embodiments, the cargo proteins may comprise four WW domains or WW domain variants from the human ITCH protein having the amino acid sequence:

PLPPGWEQRVDQHGRVYYVDHVEKRTTWDRPEPLPPGWERRVDNMGRIYYVDHFT

RTTTWQRPTLESVRNYEQWQLQRSQLQGAMQQFNQRFIYGNQDLFATSQSKEFDPL

15 GPLPPGWEKRTDSNGRVYFVNHNTTRITQWEDPRSQQQLNEKPLPEGWEMRFTVDGI

PYFVDHNRRTTTYIDPRT (SEQ ID NO: 19).

The cargo proteins, described herein, that are fused to at least one WW domain or WW domain variant are non-naturally occurring, that is, they do not exist in nature.

In some embodiments, one or more WW domains may be fused to the N-terminus of a cargo protein. In other embodiments, one or more WW domains may be fused to the C-terminus or the N-terminus of a cargo protein. In yet other embodiments, one or more WW domains may be inserted into a cargo protein. It should be appreciated that the WW domains may be configured in any number of ways to maintain function of the cargo protein, which can be tested by methods known to one of ordinary skill in the art.

25 The cargo protein of the inventive microvesicles may be a protein comprising at least one WW domain. For example, the cargo protein may be a WW domain containing protein or a protein fused to at least one WW domain. In some embodiments, the cargo protein may be a Cas9 protein or Cas9 variant fused to at least one WW domain. In some embodiments, the cargo protein may be a recombinant cargo protein. For example the recombinant cargo

30 protein may be a Cas9 protein, or Cas9 variant, fused to at least one nuclear localization sequence (NLS). A NLS, as referred to herein, is an amino acid sequence that facilitates the import of a protein into the cell nucleus by nuclear transport. In some embodiments, a NLS is fused to the N-terminus of a Cas9 protein, or Cas9 variant. In some embodiments, a NLS is fused to the C-terminus of Cas9 protein, or Cas9 variant. In some embodiments, Cas9 is

fused to at least one, at least two, at least three, at least four, at least five, at least six, at least seven, at least eight, at least nine, at least ten, or more nuclear localization sequences (NLSs). In certain embodiments, one NLS is fused to the N-terminus, and one NLS is fused to the C-terminus of the Cas9 protein to create a recombinant NLS:Cas9:NLS fusion protein. In certain embodiments, the Cas9 protein, or Cas9 variant, fused to at least one NLS may also be fused to at least one WW domain. It should be appreciated that, as described above, the WW domains may be configured in any number of ways such that the Cas9 protein or Cas9 variant may be loaded into an ARMM for delivery to a target cell and translocate into the nucleus of the target cell to perform its nuclease function. In certain embodiments, one or more WW domains are fused to the N-terminus of a recombinant NLS:Cas9:NLS fusion protein. In certain embodiments, one or more WW domains are fused to the C-terminus of a recombinant NLS:Cas9:NLS fusion protein. In certain embodiments, the cargo protein comprises the sequence (SEQ ID NO: 109) or (SEQ ID NO: 110). In certain embodiments, the cargo protein consists of the sequence (SEQ ID NO: 109) or (SEQ ID NO: 110). In certain embodiments, the cargo protein consists essentially of (SEQ ID NO: 109) or (SEQ ID NO: 110).

The following amino acid sequences are exemplary Cas9 cargo protein sequences that have either 2 WW domains (SEQ ID NO: 109) or 4 WW domains (SEQ ID NO: 110), which were cloned into the AgeI site of the pX330 plasmid (Addgene).

MPLPPGWEEQRVDQHGRVYYVDHVEKRTTWDRPEPLPPGWERRVDNMGRIYYVDHFTRTTTWQ
 RPTLTGATMDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIIHGVPAADKKYSIGLDIGTNS
 VGWAVITDEYKVP SKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKN
 RICYLQEIFSNEMAKVDDSFHRLSESLVEEDKKHERHP IFGNIVDEVAYHEKYPTIYHLR
 25 KKLVDSTDKADLRILIYLAHAMIKFRGHFLIEGDLNPDNSDVKLF IQLVQTYNQLFEEENPI
 NASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIASLGLTPNFKSNFDLAEDAK
 LQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRY
 DEHHQDLTLLKALVRQQLPPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTE
 ELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIP
 30 YYVGPLARGNSRFAMTRKSEETITPWNFEVVDKGASQSFIERMTNFDKNLPNEKVLPHK
 SLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIE
 CFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTTLTFEDREMIEER
 LKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMLI
 HDDSLTFKEDIQKAQVSGQDLSLHEHIANLAGSPAIIKKGILQTVKVVDELVKVMGRHKPENI
 35 VIEMARENQTTQKGQNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRD
 MYVDQELDINRLSDYVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKMKKNYW
 RQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRLVETRQITKHVAQILDSRMNTKYDE
 NDKLIREVKVITLKSCLVSDFRKDFQFYKVIINNYHHAHDAYLNAVVGITALIKKYPKLESE
 FVYGDYKVVYDVRKMIKSEQEI GKATAYFFYSNIMNFFKTEITLANGEIRKRPLIETNGET
 40 GEIVWDKGRDFATVRKVLSPQVNIKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKY

GGFDSPTVAYSVLVVAKEKGKSKKLKSVKELLGITIMERSSSFENPIDFLEAKGYKEVKKD
 LIIKLPKYSLEFELNGRKRMLASAGELQKGNELALPSKYVNFYLYLASHYEKLKGGSPEDNEQK
 QLFVEQHKHYLDEIIIEQISEFSKRVILADANLDKVL SAYNKHDPKIREQAENI IHLFTLTN
 LGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQ SITGLYETRIDLSQLGGDKRPAATKKAGQ
 5 AKKKK (SEQ ID NO: 109)

MPLPPGWEQRVDQHGRVYYVDHVEKRTTWDRPEPLPPGWERRVDNMGRIYYVDHFTRTTTWQ
 RPTLESVRNIEQWQLQRSQLQGAMQQFNQRFIYGNQDLFATSQSKEFDPLGPLPPGWEKRTD
 SNGRVYFVNHNTRITQWEDPRSQQQLNEKPLPEGWEMRFTVDGIPYFVDHNRRTTTYIDPRT
 10 GGGTGATMDYKDHGDYKDHIDYKDDDDKMAPKKRKVGIHGVPAAADKKYSIGLDIGTNSV
 GWAVITDEYKVP SKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNR
 ICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHP IFGNIVDEVAYHEKYPTIYHLRK
 KLVDSTDKADLRLLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEEENP IN
 ASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIASLGLTPNFKSNFDLAEDAKL
 15 QLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYD
 EHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEE
 LLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPY
 YVGPLARGNSRFAMTRKSEETITPWNFEVVVDKGASQSFIERMTNFDKNLPNEKVLPKHS
 LLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIEC
 20 FDSVEISGVEDRFNASLGTYHDLLKIKDKDFLDNEENEDILEDIVLTTLTLFEDREMIEERL
 KTYAHLFDDKVMKQLKRRRYTGWRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMLIHL
 DDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIAKKGILQTVKVDELVKVMGRHKPENIV
 IEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYLQNGRDM
 YVDQELDINRLSDYVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKMKNYWR
 25 QLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDEN
 DKLIREVKVITLKSCLVSDFRKDFQFYKVRINNYYHHAHDAYLNAVVGTAIIKKYPKLESEF
 VYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
 EIVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYG
 GFDSP TVAYSVLVVAKEKGKSKKLKSVKELLGITIMERSSSFENPIDFLEAKGYKEVKKDL
 30 IIKLPKYSLEFELNGRKRMLASAGELQKGNELALPSKYVNFYLYLASHYEKLKGGSPEDNEQKQ
 L FVEQHKHYLDEIIIEQISEFSKRVILADANLDKVL SAYNKHDPKIREQAENI IHLFTLTNL
 GAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQ SITGLYETRIDLSQLGGDKRPAATKKAGQA
 KKKK (SEQ ID NO: 119)

35 The microvesicles described herein may further comprise a nucleic acid. In some
 embodiments, the microvesicles may comprise at least one guide RNA (gRNA), which may
 be associated, for example, with a nuclease or a nickase. As one example, a gRNA may be
 associated with a Cas9 cargo protein or Cas9 variant cargo protein. The gRNA may comprise
 a nucleotide sequence that complements a target site, which mediates binding of the
 40 nuclease/RNA complex to said target site and providing the sequence specificity of the
 nuclease:RNA complex. In certain embodiments, the gRNA comprises a nucleotide
 sequence that is complementary to any target known in the art. For example, the gRNA may
 comprise a nucleotide sequence that is complementary to a therapeutic target (*e.g.*, APOC3,
 alpha 1 antitrypsin, HBV, or HIV). In certain embodiments the gRNA comprises the

sequence complementary to enhanced green fluorescent protein (EGFP). For example, the gRNA sequence may be encoded by the nucleic acid sequence set forth in SEQ ID NO: 113.

The following is an exemplary nucleic acid sequence that encodes a guide RNA (gRNA) that targets EGFP. The EGFP target sequence is underlined below.

ATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCATCCTGGTCGAG
 CTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGC
GATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTG
 10 CCCGTGCCCTGGCCACCCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCA
 GCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCCGA
 AGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGAC
 CCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAA
 GGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAA
 15 CTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAA
 GGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGA
 CCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCCGACAA
 CCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGA
 TCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGAC
 20 GAGCTGTACAAG (SEQ ID NO: 113)

TSG101

In certain embodiments, the inventive microvesicles further comprise TSG101. Tumor susceptibility gene 101, also referred to herein as TSG101, is a protein encoded by this gene belongs to a group of apparently inactive homologs of ubiquitin-conjugating enzymes. The protein contains a coiled-coil domain that interacts with stathmin, a cytosolic phosphoprotein implicated in tumorigenesis. TSG101 is a protein that comprises a UEV domain, and interacts with ARDC1. Exemplary, non-limiting TSG101 protein sequences are provided herein, and additional, suitable TSG101 protein sequences, isoforms, and variants according to aspects of this invention are known in the art. It will be appreciated by those of skill in the art that this invention is not limited in this respect. Exemplary TSG101 sequences include the following:

>gil5454140|reflNP_006283.1| tumor susceptibility gene 101 protein [*Homo sapiens*]

MAVSESQLKKMVSKYKYRDLTVRETVNVITLYKDLKPVLDSYVFNDGSSRELMNLT
GTIPVPYRGNTYNIPICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLH
EWKHPQSDLLGLIQVMIVVFGDEPPVFSRPISASYPPYQATGPPNTSYMPGMPGGISP
 5 YPSGYPPNPSGYPGCPYPPGGYPATTSSQYPSQPPVTTVGPSRDGTISED TIRASLISA
 VSDKLRWRMKEEMDRAQAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKN
 IELLKKKDEELSSALEKMENQSENNDIDEVIPTAPLYKQILNLYAEENAIEDTIFYLGE
 ALRRGVIDLDVFLKHVRLLSRKQFQLRALMQKARKTAGLSDLY (SEQ ID NO: 20)

10 >gil11230780|reflNP_068684.1| tumor susceptibility gene 101 protein [*Mus musculus*]

MAVSESQLKKMMSKYKYRDLTVRQTVNVIAMYKDLKPVLDSYVFNDGSSRELVNL
TGTIPVRYRGNIYNIPICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYL
HDWKHPRSELLELIQIMIVIFGEEPPVFSRPTVSASYPPYTATGPPNTSYMPGMPSGIS
 AYPSPGYPPNPSGYPGCPYPPAGYPATTSSQYPSQPPVTTVGPSRDGTISED TIRASLIS
 15 AVSDKLRWRMKEEMDGAQAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDK
 NIELKKKDEELSSALEKMENQSENNDIDEVIPTAPLYKQILNLYAEENAIEDTIFYLGE
 EALRRGVIDLDVFLKHVRLLSRKQFQLRALMQKARKTAGLSDLY (SEQ ID NO: 21)

>gil48374087|reflNP_853659.2| tumor susceptibility gene 101 protein [*Rattus norvegicus*]

20 MAVSESQLKKMMSKYKYRDLTVRQTVNVIAMYKDLKPVLDSYVFNDGSSRELVNL
TGTIPVRYRGNIYNIPICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYL
HDWKHPRSELLELIQIMIVIFGEEPPVFSRPTVSASYPPYTAAGPPNTSYLPSMPSGISA
 YPSGYPPNPSGYPGCPYPPAGYPATTSSQYPSQPPVTTAGPSRDGTISED TIRASLISA
 VSDKLRWRMKEEMDGAQAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKN
 25 IELLKKKDEELSSALEKMENQSENNDIDEVIPTAPLYKQILNLYAEENAIEDTIFYLGE
 ALRRGVIDLDVFLKHVRLLSRKQFQLRALMQKARKTAGLSDLY (SEQ ID NO: 22)

The UEV domain in these sequences includes amino acids 1-145 (underlined in the sequences above). The structure of UEV domains is known to those of skill in the art (see,
 30 *e.g.*, Owen Pornillos *et al.*, Structure and functional interactions of the Tsg101 UEV domain, *EMBO J.* 2002 May 15; 21(10): 2397–2406, the entire contents of which are incorporated herein by reference).

Cas9 cargo proteins fused to minimal ARRDC1

In some aspects, microvesicles, *e.g.*, ARMMs, are provided that comprise a minimal ARRDC1 protein, or variant thereof, fused to a Cas9 protein or Cas9 variant. In some aspects, fusion proteins are provided that comprise a minimal ARRDC1 protein, or variant thereof, fused to a Cas9 protein and/or a TSG101 protein, or variant thereof, fused to a Cas9 protein. In some aspects, expression constructs are provided that encode a minimal ARRDC1 protein, or variant thereof, fused to a Cas9 cargo protein and/or a TSG101 protein, or variant thereof, fused to a Cas9 cargo protein. In some embodiments, the minimal ARRDC1 protein variant is a C-terminal minimal ARRDC1 protein variant. In some embodiments, the TSG101 protein variant comprises a TSG101 UEV domain. In some embodiments, the TSG101 protein variant comprises the UEV domain and comprises at least 150, at least 160, at least 170, at least 180, at least 190, at least 200, at least 210, at least 220, at least 230, at least 240, at least 250, at least 260, at least 270, at least 280, at least 290, or at least 300 contiguous amino acids of the TSG101 sequence.

Some aspects of this invention provide ARRDC1 fusion proteins that comprise a minimal ARRDC1 protein or a variant thereof, and an endonuclease, (*e.g.*, a Cas9 protein, or Cas9 variant), associated with the minimal ARRDC1 protein or variant thereof. In some embodiments the endonuclease is covalently linked to the minimal ARRDC1 protein, or variant thereof. The endonuclease, for example, may be covalently linked to the N-terminus, the C-terminus, or within the amino acid sequence of the minimal ARRDC1 protein.

In certain embodiments, the endonuclease (*e.g.*, Cas9 protein or Cas9 variant) is fused to the C-terminus of the minimal ARRDC1 protein or protein variant, or to the C-terminus of the TSG101 protein or protein variant. The Cas9 protein or Cas9 variant may also be fused to the N terminus of the minimal ARRDC1 protein or protein variant, or to the N terminus of the TSG101 protein or protein variant. In some embodiments, the Cas9 protein or Cas9 variant may be within the minimal ARRDC1 or TSG101 protein or variants thereof.

In certain embodiments, the Cas9 protein is associated with a minimal ARRDC1 protein, a minimal ARRDC1 variant, a TSG101 protein, or a TSG101 variant via a covalent bond. In some embodiments, the Cas9 protein is associated with the minimal ARRDC1 protein, the minimal ARRDC1 protein variant, the TSG101 protein, or the TSG101 protein variant via a linker. In some embodiments, the linker is a cleavable linker, for example, the linker may contain a protease recognition site. The protease recognition site of the linker may be recognized by a protease expressed in a target cell, resulting in the Cas9 protein fused to the minimal ARRDC1 protein or variant thereof or the TSG101 protein variant thereof being released into the cytoplasm of the target cell upon uptake of the ARMM. A person

skilled in the art would appreciate that any number of linkers may be used to fuse the Cas9 protein or Cas9 variant to the minimal ARRDC1 protein or variant thereof or the TSG101 protein or variant thereof.

The Cas9 protein or Cas9 variant associated with a minimal ARRDC1 protein, a
5 minimal ARRDC1 protein variant, a TSG101 protein, or a TSG101 protein variant, may further include a nuclear localization sequence (NLS). In some embodiments, the Cas9 fusion protein is fused to at least one NLS. In some embodiments, one or more nuclear localization sequences (NLSs) are fused to the N-terminus of Cas9. In some embodiments, one or more NLSs are fused to the C-terminus of Cas9. In some embodiments, Cas9 is fused
10 to at least one, at least two, at least three, at least four, at least five, at least six, at least seven, at least eight, at least nine, at least ten, or more NLSs. It should be appreciated that one or more NLSs may be fused to Cas9 to allow translocation of Cas9 fusion protein into the nucleus of a target cell. In some embodiments, the Cas9 protein fused to at least one NLS is associated with ARRDC1, a minimal ARRDC1 protein variant, a TSG101 protein, or a
15 TSG101 protein variant via a linker. In some embodiments, the linker contains a protease recognition site. In other embodiments, the linker contains a UV-cleavable moiety. In some embodiments, the protease recognition site is recognized by a protease expressed in a target cell, resulting in the Cas9 protein fused to at least one NLS being released from the minimal ARRDC1 protein or variant thereof or the TSG101 protein or variant thereof into the
20 cytoplasm, where it may translocate into the nucleus upon uptake of the ARMM.

RNA binding proteins

Some aspects of the disclosure relate to proteins that bind to RNA. In some
embodiments, the RNA binding protein is a naturally-occurring protein, or non-naturally-
25 occurring variant thereof, or a non-naturally occurring protein that binds to an RNA, for example, an RNA with a specific sequence or structure.

In certain embodiments, the RNA binding protein is a trans-activator of transcription (Tat) protein that specifically binds a trans-activating response element (TAR element). An exemplary Tat protein comprises the amino acid sequence as set forth in SEQ ID NO: 65
30 (Table 1). Exemplary amino acid sequences of Tat proteins, as well as Tat protein fragments that bind TAR elements, are shown in Table 1. In some embodiments, the RNA binding protein is at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99% identical to the amino acid sequence of any one of SEQ ID NOs: 65-84, and binds a TAR element. In some embodiments, the RNA binding protein has at least 10, at least 15, at least 20, at least 25, at

least 30, at least 35, at least 40, at least 45, at least 50, at least 55, at least 60, at least 65, at least 70, at least 75, at least 80, at least 85, at least 90, at least 95, at least 100, at least 105, at least 110, at least 115, at least 120, at least 125, or at least 130 identical contiguous amino acids of any one of SEQ ID NOs: 65-84, and binds a TAR element. In some embodiments, the RNA binding protein has 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50 or more mutations compared to any one of the amino acid sequences set forth in SEQ ID NOs: 65-84, and binds a TAR element. In some embodiments, the RNA binding protein comprises any one of the amino acid sequences set forth in SEQ ID NOs: 65-84. In some embodiments, the Tat protein comprises an amino acid sequence as set forth in any one of SEQ ID NOs: 65-84. The RNA binding protein may also be a variant of a Tat protein that is capable of associating with a TAR element. Tat proteins, as well as variants of Tat proteins that bind to a TAR element, are known in the art and have been described previously, for example, in Kamine et al., "Mapping of HIV-1 Tat Protein Sequences Required for Binding to Tar RNA", *Virology* 182, 570-577 (1991); and Patel, "Adaptive recognition in RNA complexes with peptides and protein modules" *Curr Opin Struct Biol.* 1999 Feb;9(1):74-87; the entire contents of each of which are incorporated herein by reference. In some embodiments, the Tat protein is an HIV-1 Tat protein, or variant thereof. In some embodiments, the Tat protein is bovine immunodeficiency virus (BIV) Tat protein, or variant thereof.

A Tat protein is a nuclear transcriptional activator of viral gene expression that is essential for viral transcription from the LTR promoter and replication; it acts as a sequence-specific molecular adapter, directing components of the cellular transcription machinery to the viral RNA to promote processive transcription elongation by the RNA polymerase II (RNA pol II) complex, thereby increasing the level of full-length transcripts. Tat binds to a hairpin structure at the 5'-end of all nascent viral mRNAs referred to as the transactivation responsive RNA element (TAR RNA) in a CCNT1-independent mode.

The Tat protein consists of several domains, one is a short lysine and arginine rich region important for nuclear localization. The nine amino acid basic region of HIV-1 Tat is found at positions 49-57 of SEQ ID NO: 65, and is capable of binding a TAR element. In some embodiments, the Tat sequence comprises the nine amino acid basic region of Tat (SEQ ID NO: 73). In some embodiments the RNA binding protein comprises any one of the amino acid sequences as set forth in SEQ ID NOs: 65-67, 69, 70, or 73-84. In some embodiments, the Tat proteins are fusion proteins.

Table 1. Tat Sequences

Tat (Residue NOs)	Sequence	SEQ ID NO
HIV-1 Tat (1-101)	MEPVDPRLEPWKHPGSQPRT PCTTCYCKKC CFHCQVCFTT KALGISYGRK KRRQRRRPPQ GSQTHQVSLS KQPSSQPRGD QTGPKESSKKK VERETEADPKP	65
HIV-1 Tat (1-86)	MEPVDPRLEP WKHPGSQPRT PCTTCYCKKC CFHCQVCFTT KALGISYGRK KRRQRRRPPQ GSQTHQVSLS KQPSSQPRGD QTGPKE	66
HIV-1 Tat (37-72)	CFTT KALGISYGRK KRRQRRRPPQ GSQTHQVSLS KQ	67
HIV-1 Tat (1-45)	MEPVDPRLEP WKHPGSQPRT PCTTCYCKKC CFHCQVCFTT KALGI	68
HIV-1 Tat (49-86)	RK KRRQRRRPPQ GSQTHQVSLS KQPSSQPRGD QTGPKE	69
HIV-1 Tat (52-86)	RRQRRRPPQ GSQTHQVSLS KQPSSQPRGD QTGPKE	70
HIV-1 Tat (55-86)	RRRPPQ GSQTHQVSLS KQPSSQPRGD QTGPKE	71
HIV-1 Tat (58-86)	PPQ GSQTHQVSLS KQPSSQPRGD QTGPKE	72
HIV-1 Tat (49-57)	RK KRRQRRR	73
HIV-1 Tat (49-59)	RK KRRQRRRPP	74
HIV-1 Tat (49-61)	RK KRRQRRRPPQ G	75
HIV-1 Tat (49-63)	RK KRRQRRRPPQ GSQ	76
HIV-1 Tat (49-65)	RK KRRQRRRPPQ GSQTH	77
HIV-1 Tat (37-57)	CFTT KALGISYGRK KRRQRRR	78
HIV-1 Tat (38-62)	CFTT KALGISYGRK KRRQRRRPPQ GSQ	79
HIV-1 Tat (47-58)	GRRK KRRQRRR	80
HIV-1 Tat (46-65)	SYGRK KRRQRRRPPQ GSQTH	81
HIV-2 Tat (1-130)	METPLKAPEG SLGSYNEPSS CTSEQDAAAQ GLVSPGDEIL YQLYQPLEAC DNKCYCKKCC YHCQMCFLNK GLGIWYERKG RRRRTPKKTK AHSSASDKS ISTRTGNSQP EKKQKKTLET ALETIGGPGR	82
BIV Tat	MPGPWVAMIM LPQPKESFGG KPIGWLFWNT CKGPRRDCEPH CCCPICSWHC QLCFLQKNLG INYGSGPRRR GTRGKGRRIR RTASGGDQRR EADSQRSFTN MDQ	83
BIV Tat	SGPRPRGTRGKGRRIR	84

- In some embodiments, the RNA binding protein is a regulator of virion expression (Rev) protein (e.g., Rev from HIV-1), or variant thereof, that binds to a Rev response element (RRE). Rev proteins are known in the art and are known to the skilled artisan. For example, Rev proteins have been described in Fernandes et al., “The HIV-1 Rev response element: An RNA scaffold that directs the cooperative assembly of a homo-oligomeric ribonucleoprotein complex” *RNA Biology* 9:1, 6-11; January 2012; Cochrane et al., “The human immunodeficiency virus Rev protein is a nuclear phosphoprotein” *Virology* 171 (1):264–266, 1989; Grate et al., “Role REVersal: understanding how RRE RNA binds its peptide ligand” *Structure*. 1997 Jan 15;5(1):7-11; and Patel, “Adaptive recognition in RNA complexes with peptides and protein modules” *Curr Opin Struct Biol*. 1999 Feb;9(1):74-87; the entire

contents of each of which are incorporated herein by reference in their entirety. An exemplary Rev protein comprises the amino acid sequence as set forth in SEQ ID NOs: 93-95 (Table 3). In some embodiments, the RNA binding protein is at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99% identical to the amino acid sequence of any one of SEQ ID NOs: 93-95, and binds a Rev response element. In some embodiments, the RNA binding protein has at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 45, at least 50, at least 55, at least 60, at least 65, at least 70, at least 75, at least 80, at least 85, at least 90, at least 95, at least 100, at least 105, at least 110, or at least 115 identical contiguous amino acids of any one of SEQ ID NOs: 93-95, and binds a Rev response element. In some embodiments, the RNA binding protein has 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50 or more mutations compared to any one of the amino acid sequences set forth in SEQ ID NOs: 93-95, and binds a Rev response element. In some embodiments, the RNA binding protein comprises any one of the amino acid sequences set forth in SEQ ID NOs: 93-95. In some embodiments, the RNA binding protein comprises a variant of any one of the amino acid sequences as set forth in SEQ ID NOs: 93-95 that are capable of binding an RRE. Such variants would be apparent to the skilled artisan based on this disclosure and knowledge in the art and may be tested (e.g. for binding to an RRE) using routine methods known in the art.

In some embodiments, the RNA binding protein is a coat protein of an MS2 bacteriophage that specifically binds to an MS2 RNA. MS2 bacteriophage coat proteins that specifically bind MS2 RNAs are known in the art. For example MS2 phage coat proteins have been described in Parrott et al., "RNA aptamers for the MS2 bacteriophage coat protein and the wild-type RNA operator have similar solution behavior" *Nucl. Acids Res.* 28(2):489-497 (2000); Keryer-Bibens et al., "Tethering of proteins to RNAs by bacteriophage proteins" *Biol. Cell.* 100(2): 125-38 (2008); and Patel, "Adaptive recognition in RNA complexes with peptides and protein modules" *Curr Opin Struct Biol.* 1999 Feb;9(1):74-87; the entire contents of each are hereby incorporated by reference in their entirety. An exemplary MS2 phage coat protein comprises the amino acid sequence as set forth in SEQ ID NO: 99 (Table 4). In some embodiments, the RNA binding protein is at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99% identical to the amino acid sequence of SEQ ID NO: 99, and binds an MS2 RNA. In some embodiments, the RNA binding protein has at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 45, at least 50, at least 55, at least 60, at least 65, at least 70, at least 75, at least 80, at least 85, at least 90, at least

95, at least 100, at least 105, at least 110, or at least 115 identical contiguous amino acids of SEQ ID NO: 99, and binds an MS2 RNA. In some embodiments, the RNA binding protein has 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50 or more mutations compared to SEQ ID NO: 99, and binds an MS2 RNA. In some embodiments, the RNA binding protein comprises the amino acid sequence set forth in SEQ ID NO: 99. In some embodiments, the RNA binding protein comprises a fragment or variant of SEQ ID NO: 99 that is capable of binding to an MS2 RNA. Methods for testing whether variants or fragments of MS2 phage coat proteins bind to MS2 RNAs (e.g., SEQ ID NO: 99) can be performed using routine experimentation and would be apparent to the skilled artisan.

In some embodiments, the RNA binding protein is a P22 N protein (e.g., P22 N from bacteriophage), or variant thereof, that binds to a P22 boxB RNA. P22 N proteins are known in the art and would be apparent to the skilled artisan. For example, P22 N proteins have been described in Cai et al., "Solution structure of P22 transcriptional antitermination N peptide-boxB RNA complex" *Nat Struct Biol.* 1998 Mar;5(3):203-12; and Patel, "Adaptive recognition in RNA complexes with peptides and protein modules" *Curr Opin Struct Biol.* 1999 Feb;9(1):74-87; the entire contents of each are incorporated by reference herein. An exemplary P22 N that specifically binds to a protein P22 boxB RNA comprises the amino acid sequence NAKTRRHERRRKLAIERDTI (SEQ ID NO: 100).

In some embodiments, the RNA binding protein is a λ N protein (e.g., λ N from bacteriophage), or variant thereof, that binds to a λ boxB RNA. λ N proteins are known in the art and would be apparent to the skilled artisan. For example, λ N proteins have been described in Keryer-Bibens et al., "Tethering of proteins to RNAs by bacteriophage proteins" *Biol Cell.* 2008 Feb;100(2):125-38; Legault et al., "NMR structure of the bacteriophage lambda N peptide/boxB RNA complex: recognition of a GNRA fold by an arginine-rich motif" *Cell.* 1998 Apr 17;93(2):289-99; and Patel, "Adaptive recognition in RNA complexes with peptides and protein modules" *Curr Opin Struct Biol.* 1999 Feb;9(1):74-87; the entire contents of each are incorporated by reference herein. An exemplary λ N protein that specifically binds to a λ boxB comprises the amino acid sequence GSMDAQTRRRERRAEKQAQWKAAN (SEQ ID NO: 101).

In some embodiments, the RNA binding protein is a ϕ 21 N protein (e.g., ϕ 21 N from bacteriophage), or variant thereof, that binds to a ϕ 21 boxB RNA. ϕ 21 N proteins are known in the art and would be apparent to the skilled artisan. For example, ϕ 21 proteins have been described in Cilley et al. "Structural mimicry in the phage ϕ 21 N peptide-boxB RNA

complex.” RNA. 2003;9(6):663-676; and Patel, “Adaptive recognition in RNA complexes with peptides and protein modules” Curr Opin Struct Biol. 1999 Feb;9(1):74-87; the entire contents of each are incorporated by reference herein. An exemplary ϕ 21 N protein that specifically binds to a ϕ 21 boxB RNA comprises amino acid sequence

- 5 GTAKSRYKARRAELIAERR(SEQ ID NO: 102). The N peptide binds as an α -helix and interacts predominately with the major groove side of the 5' half of the boxB RNA stem-loop. This binding interface is defined by surface complementarity of polar and nonpolar interactions. The N peptide complexed with the exposed face of the ϕ 21 boxB loop is similar to the GNRA tetraloop-like folds of the related λ and P22 bacteriophage N peptide-boxB
10 RNA complexes.

- In some embodiments, the RNA binding protein is a HIV-1 nucleocapsid (e.g., nucleocapsid from HIV-1), or variant thereof, that binds to a SL3 ψ RNA. HIV-1 nucleocapsid proteins are known in the art and would be apparent to the skilled artisan. For example, HIV-1 nucleocapsid proteins have been described in Patel, “Adaptive recognition in
15 RNA complexes with peptides and protein modules” Curr Opin Struct Biol. 1999 Feb;9(1):74-87; the entire contents of which is incorporated by reference herein. An exemplary HIV-1 nucleocapsid that specifically binds to a SL3 ψ RNA comprises amino acid sequence:
MQKGNFRNQRKTVKCFNCGKEGHIAKNCRAPRKKGCWKC GKEGHQMKDCTERQA
20 N (SEQ ID NO: 103).

Binding RNAs

- Some aspects of the disclosure relate to RNA molecules that bind proteins. In some embodiments, the binding RNA is a naturally occurring RNA, or non-naturally occurring
25 variant thereof, or a non-naturally occurring RNA, that binds to a protein having a specific amino acid sequence or structure.

- In certain embodiments, the binding RNA is a trans-activating response element (TAR element), which is an RNA stem-loop structure that is found at the 5' ends of nascent human immunodeficiency virus-1 (HIV-1) transcripts and specifically bind to a trans-
30 activator of transcription (Tat) protein. In some embodiments, the TAR element is a bovine immunodeficiency virus (BIV) TAR. An exemplary TAR element comprises the nucleic acid sequence as set forth in SEQ ID NO: 84. Further exemplary TAR sequences can be found in Table 2; however, these sequences are not meant to be limiting and additional TAR element sequences that bind to a Tat protein, or variant thereof, are also within the scope of this

disclosure. The binding RNA may also be a variant of a TAR element that is capable of associating with the RNA binding protein, trans-activator of transcription (Tat protein), which is a regulatory protein that is involved in transcription of the viral genome. Variants of TAR elements that are capable of associating with Tat proteins would be apparent to the skilled artisan based on this disclosure and knowledge in the art, and are within the scope of this disclosure. Further, the association between a TAR variant and a Tat protein, or Tat protein variant, may be tested using routine methods. TAR elements and variants of TAR elements that bind to Tat proteins are known in the art and have been described previously, for example in Kamine et al., "Mapping of HIV-1 Tat Protein Sequences Required for Binding to Tar RNA" Virology 182, 570-577 (1991); and Patel, "Adaptive recognition in RNA complexes with peptides and protein modules" Curr Opin Struct Biol. 1999 Feb;9(1):74-87; the entire contents of each are incorporated by reference herein. In some embodiments, the binding RNA comprises the nucleic acid sequence as set forth in SEQ ID NOs: 85-90. In some embodiments, the binding RNA comprises a variant of any of the nucleic acid sequences set forth in SEQ ID NOs: 85-90 that are capable of binding to a Tat protein or variant thereof.

Without wishing to be bound by any particular theory, a TAR element is capable of forming a stable stem-loop structure (Muesing et al., 1987) in the native viral RNA. On the stem of TAR, a three nucleotide bulge, has been demonstrated to play a role in high-affinity binding of the Tat protein to the TAR element (Roy et al., 1990; Cordingley et al., 1990; Dingwall et al., 1989; Weeks et al., 1990). In the TAR element, the integrity of the stem and the initial U22 of the bulge may contribute to Tat protein binding (Roy et al., 1990b). Other sequences that may not affect the binding of the Tat protein to the TAR site play a role in trans-activation of transcription in vivo. One such region is the sequence at the loop, which is required for the binding of cellular factors that may interact with the Tat protein to mediate transactivation (Gatignol et al., 1989; Gaynor et al., 1989; Marciniak et al., 1990a; Gatignol et al., 1991).

Table 2. TAR Sequences

TAR	Sequence	SEQ ID NO
HIV- 1 TAR RNA +1-59	gggucucucugguuagaccagaucugagccugggagcucucuggcuaacuag ggaaccacug	85
Δ TAR	gggucucucugguuagaccagaucugagccugggagcucucuggcuaacuagggaa cccacug	86
HIV- 1TAR (shown in Figure 2)	gggucucucugguuagaccagaucugagccugggagcucucuggcuaacuag ggaacc	87

HIV- 1 TAR	agaucugagccugggagcucucu	88
Hybrid TAR	gcucguugagcucugggaagcuccgagc	89
BIV TAR	ucguguagcucauuagcuccga	90

In some embodiments, the binding RNA is a Rev response element (RRE), or variant thereof, that binds to a Rev protein (e.g., Rev from HIV-1). Rev response elements are known in the art and would be apparent to the skilled artisan for use in the present invention.

5 For example, Rev response elements have been described in Fernandes et al., “The HIV-1 Rev response element: An RNA scaffold that directs the cooperative assembly of a homo-oligomeric ribonucleoprotein complex.” RNA Biology 9:1, 6-11, January 2012; Cook et al., “Characterization of HIV-1 REV protein: binding stoichiometry and minimal RNA substrate.” Nucleic Acids Res. Apr 11; 19(7):1577–1583, 1991; Grate et al., “Role REVersal: understanding how RRE RNA binds its peptide ligand” Structure. 1997 Jan 15;5(1):7-11; and
10 Patel, “Adaptive recognition in RNA complexes with peptides and protein modules” Curr Opin Struct Biol. 1999 Feb;9(1):74-87; the entire contents of each are incorporated herein by reference. Any of the RRE nucleic acid sequences or any of the fragments of RRE nucleic acid sequences described in the above references may be used as binding RNAs in
15 accordance with this disclosure. Exemplary RRE nucleic acid sequences that bind Rev include, without limitation, those nucleic acid sequences set forth in SEQ ID NOs: 91 and 92 (Table 3).

In some embodiments, the Rev peptide may adopt a particular structure and several amino acids, rather than a single arginine, may participate in sequence-specific RNA
20 interactions. Without wishing to be bound by any particular theory, Rev recognition of the RRE, like Tat recognition of TAR, is due to direct binding. Binding can be tight ($K_d = 1-3$ nM) and highly specific for the RRE. As the concentration of Rev increases, progressively larger complexes with RRE RNA are formed, whereas Tat forms one-to-one complexes with TAR RNA.

25 Generally, a Rev protein may bind initially to a high affinity site and subsequently additional Rev molecules occupy lower affinity sites. RNAs that bind Rev have been described in Heaphy et al., “HIV-1 regulator of virion expression (Rev) protein binds to an RNA stem-loop structure located within the Rev-response element region” Cell, 1990. 60, 685-693; the entire contents of which is incorporated by reference herein.

30

Table 3. RRE/Rev Sequences

	Sequence	SEQ ID NO
HIV-1 RRE	ggucugggcgcagcgcaagcugacggguacaggcc	91
HIV-1 RRE aptamer	ggcuggacucguacuucggguacuggagaaacagcc	92
HIV-1 Rev	MAGRSGDSDEELIRTVRLIKLLYQSNPPPNPEGTRQARR NRRRRWRERQRQIHISERILGTYLGRSAEPVPLQLPPLE RLTLDNCNEDCGTSGTQGVGSPQILVESPTVLESGETKE	93
HIV-1 Rev peptide	TRQARRNRRRRWRERQR	94
Evolved HIV-1 RRE-binding peptide	RDRRRRGRSPSGAERRRRRAAAA	95

In some embodiments, the binding RNA is an MS2 RNA that specifically binds to a MS2 phage coat protein. Typically, the coat protein of the RNA bacteriophage MS2 binds a specific stem-loop structure in viral RNA (e.g., MS2 RNA) to accomplish encapsidation of the genome and translational repression of replicase synthesis. RNAs that specifically bind MS2 phage coat proteins are known in the art and would be apparent the skilled artisan. For example RNAs that bind MS2 phage coat proteins have been described in Parrott et al., “RNA aptamers for the MS2 bacteriophage coat protein and the wild-type RNA operator have similar solution behavior.” Nucl. Acids Res. 28(2): 489-497 (2000); Witherell et al., “Specific interaction between RNA phage coat proteins and RNA.” Prog Nucleic Acid Res Mol Biol. 1991;40:185-220; Stockley et al., “Probing sequence-specific RNA recognition by the bacteriophage MS2 coat protein.” Nucleic Acids Res. 1995 Jul 11;23(13):2512-8; Keryer-Bibens C., et al., “Tethering of proteins to RNAs by bacteriophage proteins.” Biol. Cell. 100(2): 125-38 (2008); and Patel. “Adaptive recognition in RNA complexes with peptides and protein modules.” Curr Opin Struct Biol. 1999 Feb;9(1):74-87; the entire contents of each are hereby incorporated by reference in their entirety. In some embodiments, an exemplary MS2 RNA that specifically binds to a MS2 phage coat protein comprises a nucleic acid sequence as set forth in any one of SEQ ID NOs: 96-98 (Table 4). In some embodiments, the binding RNA comprises the nucleic acid sequence of any one of SEQ ID NOs: 96, 97, or 98.

Table 4. MS2 Sequences

MS2	Sequence	SEQ ID NO
Bacteriophage MS2 RNA	acaugaggauuacccaugu	96
MS2 RNA	ccggagggaucaccacggg	97
MS2 RNA	ccacagucacuggg	98

Bacteriophage MS2 Coat Protein	ASNFTQFVLVDNGGTGDVTVAPSNFANGVAEWISSNSR SQAYKVTCSVRQSSAQNRKYTIKVEVPKVAT QTVGGVELPVAAWRSYLNMEITPIFATNSDCELIVKA MQ GLLKDGNIPI SAIAANSIGY	99
-----------------------------------	--	----

In some embodiments, the binding RNA is an RNA that specifically binds to a P22 N protein (e.g., P22 N from bacteriophage), or variant thereof. P22 N proteins are known in the art and would be apparent to the skilled artisan. For example, P22 N proteins have been described in Cai et al., "Solution structure of P22 transcriptional antitermination N peptide-boxB RNA complex" Nat Struct Biol. 1998 Mar;5(3):203-12; Weiss, "RNA-mediated signaling in transcription" Nat Struct Biol. 1998 May;5(5):329-33; and Patel, "Adaptive recognition in RNA complexes with peptides and protein modules" Curr Opin Struct Biol. 1999 Feb;9(1):74-87; the entire contents of each are incorporated by reference herein. An exemplary P22 boxB RNA that specifically binds to a P22 N protein comprises a nucleic acid sequence as set forth in ggcugacaaagcgc (SEQ ID NO: 104).

In some embodiments, the binding RNA is an RNA that specifically binds to a λ N protein (e.g., λ N from bacteriophage), or variant thereof. λ N proteins are known in the art and would be apparent to the skilled artisan. For example, λ N proteins have been described in Keryer-Bibens et al., "Tethering of proteins to RNAs by bacteriophage proteins." Biol Cell. 2008 Feb;100(2):125-38; Weiss, "RNA-mediated signaling in transcription." Nat Struct Biol. 1998 May;5(5):329-33; Legault et al., "NMR structure of the bacteriophage lambda N peptide/boxB RNA complex: recognition of a GNRA fold by an arginine-rich motif." Cell. 1998 Apr 17;93(2):289-99; and Patel, "Adaptive recognition in RNA complexes with peptides and protein modules." Curr Opin Struct Biol. 1999 Feb;9(1):74-87; the entire contents of each are incorporated by reference herein. An exemplary λ boxB RNA that specifically binds to a λ N protein comprises a nucleic acid sequence as set forth in gggcccugaagaaggccc (SEQ ID NO: 105).

In some embodiments, the binding RNA is an RNA that specifically binds to a ϕ 21 N protein (e.g., ϕ 21 N from bacteriophage), or variant thereof. ϕ 21 N proteins are known in the art and would be apparent to the skilled artisan. For example, ϕ 21 proteins have been described in Cilley et al. "Structural mimicry in the phage ϕ 21 N peptide-boxB RNA complex." RNA. 2003;9(6):663-676; and Patel, "Adaptive recognition in RNA complexes with peptides and protein modules." Curr Opin Struct Biol. 1999 Feb;9(1):74-87; the entire contents of each are incorporated by reference herein. An exemplary ϕ 21 boxB RNA that

specifically binds to a ϕ 21 N protein comprises a nucleic acid sequence as set forth in
ucucaaccuaaccguugaga (SEQ ID NO: 106).

In some embodiments, the binding RNA is an RNA that specifically binds to an HIV-
1 nucleocapsid protein (e.g., nucleocapsid from HIV-1) or variant thereof. HIV-1
5 nucleocapsid proteins are known in the art and would be apparent to the skilled artisan. For
example, HIV-1 nucleocapsid proteins have been described in Patel, "Adaptive recognition in
RNA complexes with peptides and protein modules." Curr Opin Struct Biol. 1999
Feb;9(1):74-87; the entire contents of which is incorporated by reference herein. An
exemplary SL3 ψ RNA that specifically binds to a HIV-1 nucleocapsid comprises a nucleic
10 acid sequence as set forth in ggacuaagcggagguagucc (SEQ ID NO: 107).

It should be appreciated that the binding RNAs of the present disclosure need not be
limited to naturally-occurring RNAs or non-naturally-occurring variants thereof, that have
recognized protein binding partners. In some embodiments, the binding RNA may also be a
synthetically produced RNA, for example an RNA that is designed to specifically bind to a
15 protein (e.g., an RNA binding protein). In some embodiments, the binding RNA is designed
to specifically bind to any protein of interest, for example ARRDC1. In some embodiments,
the binding RNA is an RNA produced by the systematic evolution of ligands by exponential
enrichment (SELEX). SELEX methodology would be apparent to the skilled artisan and has
been described previously, for example in U.S. Pat. Nos. 5,270,163; 5,817,785; 5,595,887;
20 5,496,938; 5,475,096; 5,861,254; 5,958,691; 5,962,219; 6,013,443; 6,030,776; 6,083,696;
6,110,900; 6,127,119; and 6,147,204; U.S. Appln 20030175703 and 20030083294, Potti et
al., Expert Opin. Biol. Ther. 4:1641-1647 (2004), and Nimjee et al., Annu. Rev. Med.
56:555-83 (2005). The technique of SELEX has been used to evolve aptamers to have
extremely high binding affinity to a variety of target proteins. See, for example, Trujillo U.
25 H., et al., "DNA and RNA aptamers: from tools for basic research towards therapeutic
applications". Comb Chem High Throughput Screen 9 (8): 619-32 (2006) for its disclosure
of using SELEX to design aptamers that bind vascular endothelial growth factor (VEGF). In
some embodiments, the binding RNA is an aptamer that specifically binds a target protein,
for example, a protein found in an ARMM (e.g., ARRDC1 or TSG101).

30

Cargo RNAs

Some aspects of the disclosure provide RNAs that are associated with, for example,
incorporated into the liquid phase of, an ARMM. In some embodiments, a cargo RNA is an
RNA molecule that can be delivered via its association with or inclusion in an ARMM to a

subject, organ, tissue, or cell. In some embodiments, the cargo RNA is to be delivered to a target cell *in vitro*, *in vivo*, or *ex vivo*. In some embodiments, the cargo RNA to be delivered is a biologically active agent, i.e., it has activity in a cell, organ, tissue, and/or subject. For instance, an RNA that, when administered to a subject, has a biological effect on that subject, or is considered to be biologically active. In certain embodiments, the cargo RNA is a messenger RNA or an RNA that expresses a protein in a cell. In certain embodiments, the cargo RNA is a small interfering RNA (siRNA) that inhibits the expression of one or more genes in a cell. In some embodiments, a cargo RNA to be delivered is a therapeutic agent, for example, an agent that has a beneficial effect on a subject when administered to a subject.

In some embodiments, the cargo RNA to be delivered to a cell is an RNA that expresses a transcription factor, a tumor suppressor, a developmental regulator, a growth factor, a metastasis suppressor, a pro-apoptotic protein, a nuclease, or a recombinase. In some embodiments, the cargo RNA to be delivered is an RNA that expresses p53, Rb (retinoblastoma protein), a BIM protein, BRCA1, BRCA2, PTEN, adenomatous polyposis coli (APC), CDKN1B, cyclin-dependent kinase inhibitor 1C, HEPACAM, INK4, Mir-145, p16, p63, p73, SDHB, SDHD, secreted frizzled-related protein 1, TCF21, TIG1, TP53, tuberous sclerosis complex tumor suppressors, Von Hippel-Lindau (VHL) tumor suppressor, CD95, ST7, ST14, a BCL-2 family protein, a caspase; BRMS1, CRSP3, DRG1, KAI1, KISS1, NM23, a TIMP-family protein, a BMP-family growth factor, EGF, EPO, FGF, G-CSF, GM-CSF, a GDF-family growth factor, HGF, HDGF, IGF, PDGF, TPO, TGF- α , TGF- β , VEGF; a zinc finger nuclease, Cre, Dre, or FLP recombinase.

In some embodiments, the cargo RNA may be an RNA that inhibits expression of one or more genes in a cell. For example, in some embodiments, the cargo RNA is a microRNA (miRNA), a small interfering RNA (siRNA), or an antisense RNA (asRNA).

In some embodiments, the cargo RNA to be delivered comprises a messenger RNA (mRNA), a ribosomal RNA (rRNA), a signal recognition particle RNA (SRP RNA), or a transfer RNA (tRNA). In some embodiments, the cargo RNA to be delivered comprises a small nuclear RNA (snRNA), a small nucleolar (snoRNA), a SmY RNA (smY), a guide RNA (gRNA), a ribonuclease P (RNase P), a ribonuclease MRP (RNase MRP), a Y RNA, a telomerase RNA component (TERC), or a spliced leader RNA (SL RNA). In some embodiments, the cargo RNA to be delivered comprises an antisense RNA (asRNA), a cis-natural antisense sequence (cis-NAT), a CRISPR RNA (crRNA), a long noncoding RNA (lncRNA), a microRNA (miRNA), a piwi-interacting RNA (piRNA), a small interfering RNA (siRNA), or a trans-acting siRNA (tasiRNA).

In some embodiments, the cargo RNA to be delivered is a diagnostic agent. In some embodiments, the cargo RNA to be delivered is a prophylactic agent. In some embodiments, the cargo RNA to be delivered is useful as an imaging agent. In some of these embodiments, the diagnostic or imaging agent is, and in others it is not, biologically active.

5 In some embodiments, any of the cargo RNAs provided herein are associated with a binding RNA. In some embodiments, the cargo RNA is covalently associated with the binding RNA. In some embodiments, the cargo RNA and the binding RNA are part of the same RNA molecule, (e.g., an RNA from a single transcript). In some embodiments, the cargo RNA and the binding RNA are covalently associated via a linker. In some
10 embodiments, the linker comprises a nucleotide or nucleic acid (e.g., DNA or RNA). In some embodiments, the linker comprises RNA. In some embodiments, the linker comprises at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 15, at least 20, at least 25, at least 30, at least 40, at least 50, at least 100, at least 150, at least 200, at least 250, at least 300, at least 400, or at least 500 nucleotides (e.g.,
15 DNA or RNA).

In other embodiments, the cargo RNA is non-covalently associated with the binding RNA. For example, the cargo RNA may associate with the binding RNA via complementary base pairing. In some embodiments, the cargo RNA is bound to the binding RNA via at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least
20 15, at least 20, at least 25, at least 30, at least 40, at least 50, complementary base pairs, which may be contiguous or non-contiguous. In some embodiments, the cargo RNA is bound to the binding RNA via at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 15, at least 20, at least 25, at least 30, at least 40, at least 50 contiguous complementary base pairs.

25 It should be appreciated that any of the RNAs provided herein (e.g., binding RNAs, cargo RNAs, and/or binding RNAs fused to cargo RNAs) may comprise one or more modified oligonucleotides. In some embodiments, any of the RNAs described herein may be modified, e.g., comprise a modified sugar moiety, a modified internucleoside linkage, a modified nucleotide and/or combinations thereof. In some embodiments, RNA
30 oligonucleotides of the invention can be stabilized against nucleolytic degradation such as by the incorporation of a modification, e.g., a nucleotide modification. For example, nucleic acid sequences of the invention include a phosphorothioate at least the first, second, or third internucleotide linkage at the 5' or 3' end of the nucleotide sequence. As another example, the nucleic acid sequence can include a 2'-modified nucleotide, e.g., a 2'-deoxy, 2'-deoxy-2'-

fluoro, 2'-O-methyl, 2'-O-methoxyethyl (2'-O-MOE), 2'-O-aminopropyl (2'-O-AP), 2'-O-dimethylaminoethyl (2'-O-DMAOE), 2'-O-dimethylaminopropyl (2'-O-DMAP), 2'-O-dimethylaminoethoxyethyl (2'-O-DMAEOE), or 2'-O--N-methylacetamido (2'-O--NMA). As another example, the nucleic acid sequence can include at least one 2'-O-methyl-modified nucleotide, and in some embodiments, all of the nucleotides include a 2'-O-methyl modification. In some embodiments, the nucleic acids are "locked," i.e., comprise nucleic acid analogues in which the ribose ring is "locked" by a methylene bridge connecting the 2'-O atom and the 4'-C atom.

Any of the modified chemistries or formats of RNA oligonucleotides described herein can be combined with each other, and that one, two, three, four, five, or more different types of modifications can be included within the same molecule.

In some embodiments, the RNA oligonucleotide may comprise at least one bridged nucleotide. In some embodiments, the oligonucleotide may comprise a bridged nucleotide, such as a locked nucleic acid (LNA) nucleotide, a constrained ethyl (cEt) nucleotide, or an ethylene bridged nucleic acid (ENA) nucleotide. Examples of such nucleotides are disclosed herein and known in the art. In some embodiments, the oligonucleotide comprises a nucleotide analog disclosed in one of the following United States Patent or Patent Application Publications: US 7,399,845, US 7,741,457, US 8,022,193, US 7,569,686, US 7,335,765, US 7,314,923, US 7,335,765, and US 7,816,333, US 20110009471, the entire contents of each of which are incorporated herein by reference for all purposes. The oligonucleotide may have one or more 2' O-methyl nucleotides. The oligonucleotide may consist entirely of 2' O-methyl nucleotides.

Expression constructs

Some aspects of this invention provide expression constructs that encode any of the minimal ARRDC1 fusion proteins, TSG101 fusion proteins, or cargo fusion proteins described herein. In some embodiments, the expression constructs described herein may further encode a guide RNA (gRNA). It should be appreciated that the gRNA may be expressed under the control of the same promoter sequence or a different promoter sequence as any of the fusion proteins described herein. In some embodiments, an expression construct encoding a gRNA may be co-expressed with any of the expression constructs described herein.

In some embodiments, the expression constructs described herein may further encode a gene product or gene products that induce or facilitate the generation of ARMMs in cells harboring such a construct. In some embodiments, the expression constructs encode a minimal ARRDC1 protein, or variant thereof, and/or a TSG101 protein, or variant thereof. In some embodiments, overexpression of either or both of these gene products in a cell increase the production of ARMMs in the cell, thus turning the cell into a microvesicle producing cell. In some embodiments, such an expression construct comprises at least one restriction or recombination site that allows in-frame cloning of a Cas9 sequence to be fused, either at the C-terminus, or at the N-terminus of the encoded minimal ARRDC1 and/or TSG101 protein or variant thereof.

In some embodiments, the expression construct comprises (a) a nucleotide sequence encoding a minimal ARRDC1 protein, or variant thereof, operably linked to a heterologous promoter, and (b) a restriction site or a recombination site positioned adjacent to the minimal ARRDC1-encoding nucleotide sequence allowing for the insertion of a nucleotide sequence encoding an additional polypeptide in frame with the ARRDC1-encoding nucleotide sequence. In some embodiments, the expression construct comprises (a) a nucleotide sequence encoding a minimal ARRDC1 protein, or variant thereof, operably linked to a heterologous promoter, and (b) a restriction site or a recombination site positioned adjacent to the minimal ARRDC1-encoding nucleotide sequence allowing for the insertion of a Cas9 or Cas9 variant sequence in frame with the minimal ARRDC1-encoding nucleotide sequence. Some aspects of this invention provide an expression construct comprising (a) a nucleotide sequence encoding a TSG101 protein, or variant thereof, operably linked to a heterologous promoter, and (b) a restriction site or a recombination site positioned adjacent to the TSG101-encoding nucleotide sequence allowing for the insertion of a Cas9 or Cas9 variant sequence in frame with the TSG101-encoding nucleotide sequence.

The expression constructs may encode a cargo protein fused to at least one WW domain. In some embodiments, the expression constructs encode a Cas9 protein, or variant thereof, fused to at least one WW domain, or variant thereof. Any of the expression constructs, described herein, may encode any WW domain or variant thereof. For example, the expression constructs may comprise any nucleotide sequence capable of encoding a WW domain or variant thereof from the poly peptide sequence (SEQ ID NO: 6); (SEQ ID NO: 7); (SEQ ID NO: 8); (SEQ ID NO: 9); (SEQ ID NO: 10); (SEQ ID NO: 11); (SEQ ID NO: 12); (SEQ ID NO: 13); (SEQ ID NO: 14); (SEQ ID NO: 18) or (SEQ ID NO: 19).

The expression constructs, described herein, may comprise any nucleic acid sequence capable of encoding a WW domain or variant thereof. For example, a nucleic acid sequence encoding a WW domain or WW domain variant may be from the human ubiquitin ligase WWP1, WWP2, Nedd4-1, Nedd4-2, Smurf1, Smurf2, ITCH, NEDL1, or NEDL2. Exemplary nucleic acid sequences of WW domain containing proteins are listed below. It should be appreciated that any of the nucleic acids encoding WW domains or WW domain variants of the exemplary proteins may be used in the invention, described herein, and are not meant to be limiting.

Human WWP1 nucleic acid sequence (uniprot.org/uniprot/Q9H0M0).

10 GAATTCGCGGCGCGCTCGACCGCTTCTGTGGCCACGGCAGATGAAACAGAAAGGCTAAAG
AGGGCTGGAGTCAGGGGACTTCTCTTCCACCAGCTTACGGTGATGATATGGCATCTGCC
AGCTCTAGCCGGGCAGGAGTGGCCCTGCCTTTTGAGAAGTCTCAGCTCACTTTGAAAGTG
GTGTCCGCAAAGCCCAAGGTGCATAATCGTCAACCTCGAATTAACCTCTACGTGGAGGTG
GCGGTGGATGGACTCCCCAGTGAGACCAAGAAGACTGGGAAGCGCATTGGGAGCTCTGAG
15 CTTCTCTGGAATGAGATCATCATTTTGAATGTACGGCACAGAGTCATTTAGATTTAAAG
GTCTGGAGCTGCCATACCTTGAGAAATGAACTGCTAGGCACCGCATCTGTCAACCTCTCC
AACGTCTTGAAGAACAATGGGGGCAAAATGGAGAACATGCAGCTGACCTGAACCTGCAG
ACGGAGAACAAGGCAGCGTTGTCTCAGGCGGAAACTGACAATTTCTGGACGGGCCA
ACTGTTGATCTGGGAAATGTGCCTAATGGCAGTGCCCTGACAGATGGATCACAGCTGCCT
20 TCGAGAGACTCCAGTGGAACAGCAGTAGCTCCAGAGAACCGGCACCAGCCCCCAGCACA
AACTGCTTTGGTGGAAGATCCCGGACGCACAGACATTGGGTGCTTCAGCCAGAACAACC
CCAGCAACCGCGAGCAAAGCCCCGGTGCTCGGAGCCGGCACCAGCCAGCCGCTCAAGAAC
TCAGGCCACAGTGGCTTGGCCAATGGCACAGTGAATGATGAACCCACAACAGCCACTGAT
CCCGAAGAACCTTCCGTTTGGTGTGACGTCCCCACCTGCTGCACCCTTGAGTGTGACC
25 CCGAATCCCAACACGACTTCTCTCCCTGCCCCAGCCACACCGGCTGAAGGAGAGGAACCC
AGCACTTCCGGTACACAGCAGCTCCCAGCGGCTGCCAGGCCCCGACGCTCTGCCTGCT
GGATGGGAACAGCGAGAGCTGCCAACGGACGTGTCTATTATGTTGACCACAATACCAAG
ACCACCACCTGGGAGCGGCCCTTCTCCAGGCTGGGAAAAACGCACAGATCCCCGAGGC
AGGTTTTACTATGTGGATCACAATACTCGGACCACCACCTGGCAGCGTCCGACCAGCGGAG
30 TAGCTGCGCAACTATGAGCAGTGGCAGTGCAGCGGAATCAGCTCCAGGGGGCCATGCAG
CACTTCAGCCAAAGATTCTTATACCAGTTTGGAGTGTCTCGACTGACCATGATCCCCTG
GGCCCCCTCCCTCCTGTTGGGAGAAAAGACAGGACAATGGACGGGTGTATTACGTGAAC
CATAACACTCGCACGACCCAGTGGGAGGATCCCCGAGCCAGGGGATGATCCAGGAACCA
GCTTTGCCCCCAGGATGGGAGATGAAATACACCAGCGAGGGGGTGCAGATACTTTGTGGAC
35 CACAATACCCGACCAACACCTTTAAGGATCCTCGCCCGGGGTTTGAGTCCGGGACGAAG
CAAGGTTCCCTTGGTGCTTATGACCGCAGTTTTCGGTGAAGTATCACCAGTTCCGTTTC
CTCTGCCATTCAAATGCCCTACCTAGCCACGTGAAGATCAGCGTTTCCAGGCAGACGCTT
TTCAAGATTCTTCCAACAGATCATGAACATGAAACCCTATGACCTGCGCCGCGGCTT
TACATCATCATGCGTGGCGAGGAGGGCTGGACTATGGGGGCATCGCCAGAGAGTGGTTT
40 TTCTCTCTGCTCAGGAGTGCTCAACCTATGTATTGTTATTTGAATATGCCGAAAG
AACAATTACTGCCTGCAGATCAACCCGCTCTCCATCAACCCGACCACTCACCTAC
TTTCGCTTTTATAGGCAGATTATCGCCATGGCGCTGTACCATGAAAGTTTCATCGACACG
GGCTTACCCCTCCCTTTCTACAAGCGGATGCTCAATAAGAGACCAACCCGTAAAGACCTG
GAGTCCATTGACCCTGAGTTCTACAACCTCATTGTCTGGATCAAAGAGAACAACCTGGAA
45 GAATGTGGCTGGAGCTGTACTTCATCCAGGACATGGAGATACTGGCAAGGTGACGACC
CACGAGCTGAAGGAGGGCGGAGAGCATCCGGGTACGGAGGAGAAACAAGGAAGAGTAC
ATCATGCTGCTGACTGACTGGCGTTTACCCGAGGCGTGAAGAGCAGACCAAGCCTTC
CTGGATGGCTTCAACGAGGTGGCCCCGCTGGAGTGGCTGCGCTACTTTGACGAGAAAGAG
CTGGAGCTGATGCTGTGCGCATGCAGGAGATAGACATGAGCGACTGGCAGAAGAGCACC
50 ATCTACCGGCACTACACCAAGAACAGCAAGCAGATCCAGTGGTTCTGGCAGGTGGTGAAG
GAGATGGACAACGAGAAGAGGATCCGGTGCTGCAGTTGTACCCGTACCTGCCCGCTG
CCCGTCGGGGGATTTGCCGAACATCCGTAGCAACGACCACAGAAGTTTTGCATTGAC
AAAGTTGGCAAGGAACTGGCTGCCGAGAAGCCACACCTGCTTCAACCGTCTGGATCTT
CCACCTACAAGAGCTACGAACAGCTGAGAGAGAAGCTGCTGTATGCCATTGAGGAGACC

GAGGGCTTTGGACAGGAGTAACCGAGGCCGCCCTCCACGCCCCCAGCGCACATGTAG
TCCTGAGTCCCTCCCTGCCTGAGAGGCCACTGGCCCCGAGCCCTTGGGAGGCCCCCGTGG
ATGTGGCCCTGTGTGGGACCACACTGTCTCTCGCTGCTGGCAGAAAAGCCTGATCCCAG
GAGGCCCTGCAGTTCCCCCGACCCGCGGATGGCAGTCTGGAATAAAGCCCCCTAGTTGCC
5 TTTGGCCCCACCTTTGCAAAGTTCAGAGGGCTGACCCTCTCTGCAAAACTCTCCCCTGT
CCTCTAGACCCCCACCCTGGGTGTATGTGAGTGTGCAAGGGAAGGTGTGCATCCCCAGGG
GCTGCCGCGAGAGGCCGAGACCTCCTGGACTAGTTCGGCGAGGAGACTGGCCACTGGGGG
TGGCTGTTCCGGACTGAGAGCGCCAAGGGTCTTTGCCAGCAAAGGAGGTTCTGCCTGTAA
TTGAGCCTCTCTGATGATGAGATGAAGTGAAGGTCTGAGGGACGGGCCCTGGGGCTAGG
10 CCATCTCTGCCTGCCCTCCCTAGCAGGCCGACGCGTGGAGGCTGAGTCGCAGGACACATG
CCGGCCAGTTAATTCATTCTCAGCAAAATGAAGGTTTGTCTAAGCTGCCTGGGTATCCACG
GGACAAAAACAGCAAACTCCCTCCAGACTTTGTCCATGTTATAAACTTGAAAGTTGGTTG
TTGTTTGTAGGTTTGCCAGGTTTTTTTGTTTACGCCCTGCTGTCACTTTCCTGTC
(SEQ ID NO: 23)

15

Human WWP2 nucleic acid sequence (uniprot.org/uniprot/ O00308).

GAATTCGCGGCGCGCTCGACCGCTTCTGTGGCCACGGCAGATGAAACAGAAAGGCTAAAG
AGGGCTGGAGTCAGGGGACTTCTCTTCCACCAGCTTCACGGTGATGATATGGCATCTGCC
AGCTCTAGCCGGGCGAGGAGTGGCCCTGCCCTTTGAGAAGTCTCAGCTCACTTTGAAAGTG
20 GTGTCCGCAAAAGCCCAAGGTGCATAATCGTCAACCTCGAATTAACCTCCTACGTGGAGGTG
GCGGTGGATGGACTCCCCAGTGAGACCAAGAAGACTGGGAAGCGCATTGGGAGCTCTGAG
CTTCTCTGGAATGAGATCATCATTTTTGAATGTACGGCACAGAGTCATTTAGATTTAAAG
GTCTGGAGCTGCCATACCTTGAGAAATGAAGTGTCTAGGCACCGCATCTGTCAACCTCTCC
AACGTCTTGAGAACAATGGGGGCAAAATGGAGAACATGCAGCTGACCCTGAACCTGCAG
25 ACGGAGAACAAGGCAGCGTTGTCTCAGGCGGAAAACAGCAATTTTCTGGACGGGCCA
ACTGTTGATCTGGGAAATGTGCCTAATGGCAGTGCCTGACAGATGGATCACAGCTGCCT
TCGAGAGACTCCAGTGGAAACAGCAGTAGCTCCAGAGAACCAGGACACAGCCCCCAGCACA
AACTGCTTTGGTGGAAGATCCCGGACGCACAGACATTCGGGTGCTTCAGCCAGAACAACC
CCAGCAACCGGCGAGCAAAGCCCCGGTGTCTGGAGCCGGCACCAGCCAGCCGTCAGAAC
30 TCAGGCCACAGTGGCTTGGCCAATGGCACAGTGAATGATGAACCCACAACAGCCACTGAT
CCCGAAGAACCTTCCGTTGTTGGTGTGACGTCCCCACCTGCTGCACCCTTGAGTGTGACC
CCGAATCCCAACAGACTTCTCTCCCTGCCCCAGCCACACCGGCTGAAGGAGAGGAACCC
AGCACTTCGGGTACACAGCAGCTCCAGCGGCTGCCAGGCCCGCCGACGCTCTGCCTGCT
GGATGGGAACAGCGAGAGCTGCCCAACGGACGTGTCTATTATGTTGACCACAATACCAAG
35 ACCACCACCTGGGAGCGGCCCTTCTCCAGGCTGGGAAAAACGCACAGATCCCCGAGGC
AGGTTTTACTATGTGGATCACAATACTCGGACCACCACCTGGCAGCTCCGACCCGCGGAG
TACGTGCGCAACTATGAGCAGTGGCAGTCGCAGCGGAATCAGCTCCAGGGGGCCATGCAG
CACTTCAGCCAAAGATTTCTATACCAGTTTGGAGTGTCTCGACTGACCATGATCCCTTG
GGCCCCCTCCCTCCTGTTGGGAGAAAAGACAGGACAATGGACGGGTGATTACGTGAAC
40 CATAACACTCGCACGACCCAGTGGGAGGATCCCCGACCCAGGGGATGATCCAGGAACCA
GCTTTGCCCCAGGATGGGAGATGAAATACACCAGCGAGGGGGTGCATACTTTGTGGAC
CACAATACCCGCACCACCACCTTTAAGGATCCTCGCCCGGGGTTTGAGTCGGGGACGAAG
CAAGGTTCCCTGGTGTATGACCGCAGTTTTCGGTGGAAGTATCACCAGTTCCGTTTC
CTCTGCCATTCAAATGCCCTTACCTAGCCACGTGAAGATCAGCGTTTCCAGGCAGACGCTT
45 TTGGAAGATTCCCTTCAAACAGATCATGAACATGAAACCTATGACCTGCGCCGCGCGCTT
TACATCATCATGCGTGGCGAGGAGGGCCTGGACTATGGGGGCATCGCCAGAGAGTGGTTT
TTCCTCCTGTCTCAGAGGTGCTCAACCTATGTATTGTTTATTTGAATATGCCGGAAG
AACAATTACTGCCTGCAGATCAACCCCGCTCCTCCATCAACCCGAGACACCTCACCTAC
TTTCGCTTTATAGGCAGATTTCATCGCCATGGCGCTGTACCATGGAAGTTTCATCGACACG
50 GGCTTCACCCCTCCCTTTCTACAAGCGGATGCTCAATAAGAGACCAACCTGAAAGACCTG
GAGTCCATTGACCCTGAGTTCTACAACCTCATTGTCTGGATCAAAGAGAACAACCTGGAA
GAATGTGGCTGGAGCTGTACTTCATCCAGGACATGGAGATACTGGGCAAGGTGACGACC
CAGGAGCTGAAGGAGGGCGCGAGAGCATCCGGGTACGGAGGAGAGAACAAGGAAGAGTAC
ATCATGCTGCTGACTGACTGGCGTTTCACCCGAGGCGTGGAAGAGCAGACCAAAGCCTTC
55 CTGGATGGCTTCAACGAGGTGGCCCCGCTGGAGTGGCTGCGCTACTTTGACGAGAAAGAG
CTGGAGCTGATGCTGTGCGGCATGCAGGAGATAGACATGAGCGACTGGCAGAAGAGCACC
ATCTACCGGCACTACACCAAGAACAGCAAGCAGATCCAGTGGTTCTGGCAGGTGGTGAAG
GAGATGGACAACGAGAAGAGGATCCGGCTGCTGCAGTTTGTACCCGCTACCTGCCCGCTG
CCCGTCGGGGGATTTGCCGAACCTCATCGGTAGCAACGACCACAGAAGTTTTGCATTGAC

AAAGTTGGCAAGGAAACCTGGCTGCCCAGAAGCCACACCTGCTTCAACCGTCTGGATCTT
 CCACCCCTACAAGAGCTACGAACAGCTGAGAGAGAAGCTGCTGTATGCCATTGAGGAGACC
 GAGGGCTTTTGACAGGAGTAACCGAGGCCGCCCTCCACGCCCCCAGCGCACATGTAG
 TCCTGAGTCTCCCTGCCTGAGAGGCCACTGGCCCCGAGCCCTTGGGAGGCCCCCGTGG
 5 ATGTGGCCCTGTGTGGGACCACACTGTCATCTCGCTGCTGGCAGAAAAGCCTGATCCCAG
 GAGGCCCTGCAGTTCCTCCCGACCCGCGGATGGCAGTCTGGAATAAAGCCCCCTAGTTGCC
 TTTGGCCCCACCTTTGCAAAGTTCAGAGGGCTGACCTCTCTGCAAACTCTCCCCTGT
 CCTCTAGACCCACCCCTGGGTGTATGTGAGTGTGCAAGGGAAGGTGTTCATCCCCAGGG
 GCTGCCCGCAGAGGCCGGAGACCTCCTGGACTAGTTCGGCGAGGAGACTGGCCACTGGGGG
 10 TGGCTGTTCCGGGACTGAGAGCGCCAAGGGTCTTTGCCAGCAAAGGAGGTTCTGCCGTGTA
 TTGAGCCTCTCTGATGATGGAGATGAAGTGAAGGTCTGAGGGACGGGCCCTGGGGCTAGG
 CCATCTCTGCCTGCCTCCCTAGCAGGCCAGCGGTGGAGGCTGAGTCGCAGGACACATG
 CCGGCCAGTTAATTCATTCTCAGCAAATGAAGGTTTGTCTAAGCTGCCTGGGTATCCACG
 GGACAAAAACAGCAAACCTCCCTCCAGACTTTGTCCATGTTATAAACTTGAAAGTTGGTTG
 15 TTGTTTGTAGGTTTGCCAGGTTTTTTTGTTCACGCTGCTGTCACTTTCCTGTG
 (SEQ ID NO: 24)

Human Nedd4-1 nucleic acid sequence (uniprot.org/uniprot/ P46934).

ACAGTTGCCTGCCCTGGGCGGGGGCGAGCGCTCCGGTTTGTCTGGAAGCGTTCCGAAATG
 20 GCAACTTGC CGGTGGAGGTGTTCCGGCTCCTGGAGGACGAGGAAAATTCACGAATTGTG
 AGAGTAAGAGTTATAGCCGAATAGGCCTTGCCAAGAAGGATATATTTGGGAGCTAGTGAT
 CCTTACGTGAGAGTGACGTTATATGACCCAATGAATGGAGTTCTTACAAGTGTGCAAAACA
 AAAACCATTAAAAAGAGTTTGAATCCAAAGTGAAGTGAAGAAATATTATTCAGAGTTCAT
 CCTCAGCAGCACCGGCTTCTTTTGAAGTGTGTTGACGAAAACCGATTGACAAGAGATGAT
 25 TTCTAGGTCAAGTGGATGTTCCACTTTATCCATTACCGACAGAAAAATCCAAGATTGGAG
 AGACCATATACATTTAAGGATTTTGTCTTCTCATCCAAGAAGTCACAAATCAAGAGTTAAA
 GTTTATCTGAGACTAAAAATGACTTATTTACCTAAAACAGTGGCTCAGAAGATGATAAT
 GCAGAACAGGCTGAGGAATTAGAGCCTGGCTGGGTGTTTTGGACCAACCAGATGCTGCT
 TGCCATTTGCAGCAACAACAAGAACCTTCTCTCTACCTCCAGGGTGGGAAGAGAGGCAG
 30 GATATCCTTGAAGGACCTATTATGTAACCATGAATCTAGAAGAACACAGTGGAAAAGA
 CCAACCCCTCAGGACAACCTAACAGATGCTGAGAATGGCAACATTCAACTGCAAGCACAA
 CGTGCATTTACCACCAGGCGGCAGATATCCGAGGAAACAGAAAGTGTGACAACCAAGAG
 TCTTCCGAGAACTGGGAAATTATAAGAGAAGATGAAGCCACCATTGTATAGCAGCCAGGCC
 TTCCCATCACCTCCACCGTCAAGTAACCTGGATGTTCCAACCTCATCTTGCAAGAAGATTG
 35 AATGCCAGACTCACCATTTTTTGGAAATTCAGCCGTGAGCCAGCCAGCATCGAGCTCAAAT
 CATTCAGCAGAGAAGAGGCAGCTTACAAGCCTATACCTTTGAGGAACAACCTACACTTCCT
 GTGCTTTTGCTTACTTCACTCTGGATTACCACCAGGTGGGAAGAAAAACAAGATGAAAGA
 GGAAGATCATATTATGTAGATCACAATTCAGAACGACTACTTGGACAAAGCCCACTGTA
 CAGGCCACAGTGGAGACCAGTCACTGACCTCAAGCCAGAGTTCTGCAGGCCCTCAATCA
 40 CAAGCCTCCACCAGTGATTGAGGCCAGCAGGTGACCCAGCCATCTGAAATTGAGCAAGGA
 TTCCTTCTTAAAGGCTGGGAAGTCCGGCATGCACCAAATGGGAGGCCCTTCTTTATTGAC
 CACAACACTAAAACACCACCTGGGAAGATCCAAGATTGAAAATTCAGCCCATCTGAGA
 GGAAAGACATCACTTGATACTTCCAATGATCTAGGGCCTTTACCTCCAGGATGGGAAGAG
 AGAACTCACACAGATGGAAGAATCTTCTACATAAATCACAATATAAAAAGAACAATGG
 45 GAAGATCCTCGGTTGGAGAATGTAGCAATAACTGGACCAGCAGTGCCTACTCCAGGGAT
 TACAAAAGAAAGTATGAGTTCTTCCGAAGAAAGTTGAAGAAGCAGAATGACATTCCAAAC
 AAATTTGAAATGAAACTTCGCCGAGCAACTGTTCTTGAAGACTCTTACCGGAGAATTATG
 GGTGTCAAGAGAGCAGACTTCCTGAAGGCTCGACTGTGGATTGAGTTTGTGTTGAAAG
 GGATTGGATTATGGAGGAGTTGCCAGAGAATGGTTCTTCTGATCTCAAAGGAAATGTTT
 50 AACCTTATTTATGGGTTGTTGAATATTCTGCTACGGGACAATTATACCCTACAGATAAAT
 CCAAACCTCTGGATTGTGTACGAAGATCACCTCTCTTACTTCAAGTTTATTGGTCGGGTA
 GCTGGAATGGCAGTTTATCATGGCAAACCTGTTGGATGGTTTTTTCATCCGCCCATTTTAC
 AAGATGATGCTTCACAAACCAATAACCTTCATGATATGGAATCTGTGGATAGTGAATAT
 TACAATTCCTTAAGATGGATTCTTGAAAAATGACCCAACAGAATTGGACCTCAGGTTTATC
 55 ATAGATGAAGAATTTTTGGACAGACATCAACATGAGCTGAAAAATGGTGGATCAGAA
 ATAGTTGTCCAAATAAGAACAACAAAGGAATATATTATCTTGTAAATACAATGGCGATT
 GTAAACCGAATCCAGAAGCAAATGGCTGCTTTTAAAGAGGGATTCTTTGAACTAATACCA
 CAGGATCTCATCAAATTTTTGATGAAAATGAACTAGAGCTTCTTATGTGTGGACCGGGA
 GATGTTGATGTGAATGACTGGAGGGAACATACAAAGTATAAAAAATGGCTACAGTGCAAAAT

CATCAGGTTTATACAGTGGTTTTGGAAGGCTGTTTTAATGATGGATTGAGAAAAAGAATA
AGATTACTTCAGTTTGTCTACTGGCACATCTCGGGTGCCTATGAATGGATTGCTGAACATA
TACGGTTCAAATGGACCACAGTCATTTACAGTTGAACAGTGGGGTACTCCTGAAAAGCTG
CCAAGAGCTCATACCTGTTTTAATCGCCTGGACTTGCCACCTTATGAATCATTGGAAGAA
5 TTATGGGATAAACTTCAGATGGCAATTGAAAACACCCAGGGCTTTGATGGAGTTGATTAG
ATTACAAATAACAATCTGTAGTGTTTTTACTGCCATAGTTTTATAACCAAAATCTTGACT
TAAAATTTTCCGGGAACTACTAAAATGTGGCCACTGAGTCTTCCCAGATCTTGAAGAAA
ATCATATAAAAAGCATTGGAAGAAATAGTACGAC

(SEQ ID NO: 25)

10

Human Nedd4-2 nucleic acid sequence (>gil345478679|reflNM_015277.5| Homo
sapiens neural precursor cell expressed, developmentally down-regulated 4-like, E3 ubiquitin
protein ligase (NEDD4L), transcript variant d, mRNA).

15 ATGGCGACCGGGCTCGGGGAGCCGGTCTATGGACTTTCCGAAGACGAGGGAGAGTCCCGTATTCTCA
GAGTAAAAGTTGTTTTCTGGAATTGATCTCGCCAAAAAGGACATCTTTGGAGCCAGTGATCCGTATGTGAA
ACTTTCATTGTACGTAGCGGATGAGAATAGAGAACTTGCTTTGGTCCAGACAAAAACAATTAAGGACA
CTGAACCCAAAATGGAATGAAGAATTTTATTTTCAAGGTAAACCCATCTAATCACAGACTCCTATTGGAAG
TATTTGACGAAAATAGACTGACACGAGACGACTTCCCTGGGCCAGGTGGACGTGCCCTTAGTCACCTTCC
GACAGAAGATCCAACCATGGAGCGACCTATACATTTAAGGACTTTCTCCTCAGACCAAGAAGTCATAAG
20 TCTCGAGTTAAGGGATTTTTGCGATTGAAAATGGCCTATATGCCAAAAATGGAGGTCAAGATGAAGAAA
ACAGTGACCAGAGGGATGACATGGAGCATGGATGGGAAGTTGTTGACTCAAATGACTCGGCTTCTCAGCA
CCAAGAGGAACTTCTCTCTCTCTCTGCTCCCGGTGGGAAGAAAAAGTGGACAATTTAGGCCGAACCT
TACTATGTCAACCACAACAACCGGACCCTCAGTGGCACAGACCAAGCCTGATGGACGTGCTCTCGGAGT
CGGACAATAACATCAGACAGATCAACCAGGAGGCAGCACACCGGCGCTTCCGCTCCCGCAGGCACATCAG
25 CGAAGACTTGGAGCCCGAGCCCTCGGAGGGCGGGGATGTCCCGAGCCTTGGGAGACCATTTGAGAGGAA
GTGAATATCGTGGAGACTCTCTCGGTCTGGCTCTGCCCCACCACCGGCCTCCCGAGGATCTCGGACCA
GCCCTCAGGAGCTGTGAGGAACTAAGCAGAAGGCTTCAGATCACTCCAGACTCCAATGGGGAACAGTT
CAGCTCTTTGATTCAAAGAGAACCCCTCCTCAAGGTTGAGGTGATGCAGTGTACCGACGAGTTCAGAA
CAGGGCCATCTACCACCGCCATCAGTGGCCTATGTACATACCACGCGGGTCTGCCTTCAGGCTGGGAAG
30 AAAGAAAAGATGCTAAGGGGCGCACATACTATGTCAATCATAACAATCGAACCACAACCTTGGACTCGACC
TATCATGCAGCTTGCAGAAGATGGTGGCTCCGGATCAGCCACAAACAGTAACAACCATCTAATCGAGCCT
CAGATCCGCCCGGCTCGTAGCCTCAGCTCGCCAACAGTAACCTTTATCTGCCCGCTGGAGGGTGCCAAGG
ACTCACCCGTACGTGGGCTGTGAAAGACACCCCTTTCCAACCCACAGTCCCCACAGCCATCACCTTACAA
CTCCCCCAACCAACAACAAGTCAACACAGAGCTTCTTGCCACCCGGCTGGGAAATGAGGATAGCGCCA
35 AACGGCCGGCCCTTCTTCAATTGATCATAACACAAAGACTACAACCTGGGAAGATCCACGTTTGAAATTC
CAGTACATATGCGGTCAAAGACATCTTTAAACCCCAATGACCTTGGCCCCCTTCTCCTGGCTGGGAAGA
AAGAATTCACCTTGGATGGCCGAACGTTTTATATTGATCATAATAGCAAAATTACTCAGTGGGAAGACCCA
AGACTGCAGAACCCAGCTATTACTGGTCCGGCTGTCCCTTACTCCAGAGAATTTAAGCAGAAATATGACT
ACTTCAGGAAGAAATTAAGAAACCTGCTGATATCCCCAATAGGTTTGAAATGAAACTTCACAGAAATAA
40 CATATTTGAAGAGTCTATCGGAGAATTATGTCCGTGAAAAGACCAGATGTCTTAAAGCTAGACTGTGG
ATTGAGTTTGAATCAGAGAAAGGTCTTGACTATGGGGTGTGGCCAGAGAATGGTTCTTCTTACTGTCCA
AAGAGATGTTCAACCCCTACTACGGCCTCTTTGAGTACTCTGCCACGGACAACCTACACCTTCAGATCAA
CCCTAATTCAGGCCTCTGTAATGAGGATCATTGTCTCTACTTCACTTTTATTGGAAGAGTTGCTGGTCTG
GCCGTATTTTCATGGGAAGCTCTTAGATGGTTCTTCATTAGACCATTTTACAAGATGATGTTGGGAAAGC
45 AGATAACCCGTAATGACATGGAATCTGTGGATAGTGAATATTACAACCTCTTTGAAATGGATCCTGGAGAA
TGACCACTACTGAGCTGGACCTCATGTTCTGCATAGACGAAGAAACCTTTGGACAGACATATCAAGTGGAT
TTGAAGCCCAATGGGTGAGAAATAATGGTCACAAATGAAAACAAAAGGGAATATATCGACTTAGTCATCC
AGTGGAGATTTGTGAACAGGGTCCAGAAGCAGATGAACGCCTTCTTGGAGGGATTACAGAACTACTTCC
TATTGATTTGATTAATAATTTTGTGATGAAAATGAGCTGGAGTTGCTCATGTGCGGCCTCGGTGATGTGGAT
50 GTGAATGACTGGAGACAGCATTCTATTTACAAGAACGGCTACTGCCCAAACACCCCGTCATTGAGTGGT
TCTGGAAGGCTGTGCTACTCATGGACGCCGAAAAGCGTATCCGGTTACTGCAGTTTGTACAGGGACATC
GCGAGTACCTATGAATGGATTGCGCAACTTTATGGTTCCAATGGTCTCAGCTGTTTACAATAGAGCAA
TGGGGCAGTCTGAGAAACTGCCAGAGCTCACACATGCTTTAATCGCCTTGACTTACCTCCATATGAAA
CCTTTGAAGATTTACGAGAGAACTTCTCATGGCCGTGGAAAATGCTCAAGGATTTGAAGGGGTGGATTA
55 A (SEQ ID NO: 26)

Human Smurf1 nucleic acid sequence (uniprot.org/uniprot/ Q9HCE7).

ATGTCGAACCCCGGGACACGCAGGAACGGCTCCAGCATCAAGATCCGTCTGACAGTGTTA
TGTGCCAAGAACCTTGCAAAGAAAGACTTCTTCAGGCTCCCTGACCTTTTGCAAAGATT
GTGCTGGATGGGTCTGGGCAGTGCCACTCAACCGACACTGTGAAAAACACATTGGACCCA
5 AAGTGAACACAGCACTATGATCTATATGTTGGGAAAACGGATTGATAACCATTAGCGTG
TGAACCATAGAAAATTACAAAGAAACAGGGAGCTGGCTTCTGGGCTGTGTGCGGCTG
CTCTCCAATGCCATCAGCAGATTAAAAGATACCGGATACCAGCGTTTGGATCTATGCAAA
CTAAACCCCTCAGATACTGATGCAGTTCGTGGCCAGATAGTGGTCAGTTTACAGACACGA
GACAGAATAGGAACCGGCGGCTCGGTGGTGGACTGCAGAGGACTGTTAGAAAATGAAGGA
10 ACGGTGTATGAAGACTCCGGGCTGGGAGGCCGCTCAGCTGCTTCATGGAGGAACAGCC
CCTTACACAGATAGCACCGGTGCTGCTGCTGGAGGAGGGAATTGCAGGTTCTGTGGAGTCC
CCAAGTCAAGATCAAAGACTTCAGGCACAGCGGCTTCGAAACCTGATGTGCGAGGTTCA
CTACAGACGCCCCAGAACCGACACACGCCACCAGTCCCCGGAAGTGCCTGAAGGCTAC
GAACAAAGAACACAGTCCAGGGCCAAGTTTACTTTTTGCATACACAGACTGGAGTTAGC
15 ACGTGGCAGACCCCCAGGATACCAAGTCCCTCGGGGACCATTCCTGGGGGAGATGCAGCT
TTTCTATACGAATTCCTTACAAAGGCCATACATCTGAGCCAGAGACCTTAACAGTGTG
AACTGTGATGAACCTGGACCACTGCCGCCAGGCTGGGAAGTCAGAAGTACAGTTTCTGGG
AGGATATATTTGTAGATCATAATAACCGAACAACCCAGTTTACAGACCCAAGGTTACAC
CACATCATGAATCACCAGTGCCAACTCAAGGAGCCAGCCAGCCGCTGCCACTGCCAGT
20 GAGGGCTCTCTGGAGGACGAGGAGCTTCTGCCAGAGATACGAAAGAGATCTAGTCCAG
AAGCTGAAAGTCTCTCAGACACGAACTGTCGCTTCAGCAGCCCCAAGCTGGTCATTGCCGC
ATCGAAGTGTCCAGAGAAGAAATCTTTGAGGAGTCTTACCGCCAGATAATGAAGATGCGA
CCGAAAGACTTGAAAAAACGGCTGATGGTGAAATTCCTGCGGGAAGAAGGTTTGGATTAC
GGTGGTGTGGCCAGGGAGTGGCTTTACTTGTGTGCCATGAAATGCTGAATCCTTATTAC
25 GGGCTCTTCCAGTATTCTACGGACAATATTTACATGTTGCAAATAAATCCGGATTCTTCA
ATCAACCCCGACCACTTGTCTTATTTCCACTTTGTGGGGCGGATCATGGGGCTGGCTGTG
TTCCATGGACACTACATCAACGGGGGCTTCACAGTGCCTTCTACAAGCAGCTGCTGGGG
AAGCCCATCCAGCTCTCAGATCTGGAATCTGTGGACCCAGAGCTGCATAAGAGCTTGGTG
TGGATCCTAGAGAACGACATCACGCCTGTACTGGACCACACCTTCTGCGTGGAACACAAC
30 GCCTTCGGGCGGATCCTGCAGCATGAACGAAACCCAATGGCAGAAATGTGCCAGTCACA
GAGGAGAATAGAAAGAAATACGTCCGGTGTATGTAAGTGGAGGTTATGAGAGGAATC
GAAGCCCAGTCTTAGCTCTGCAGAAGGGGTTCAATGAGCTCATCCCTCAACATCTGCTG
AAGCCTTTTGACCAGAAGGAAGTGGAGCTGATCATAGCGCGCTGGATAAAATAGACTTG
AACGCTTGGAGTCCGAACACGCGGCTGAAGCACTGTGTGGCCGACAGCAACATCGTGGG
35 TGGTTCTGGCAAGCGGTGGAGACGTTTCGATGAAGAAAGGAGGGCCAGGCTCCTGCAGTTT
GTGACTGGGTCCACGCGAGTCCCGCTCCAAGGCTTCAAGGCTTTGCAAGGTTCTACAGGC
GCGGCAGGGCCCCGGCTGTTACCATCCACCTGATAGACGCGAACACAGACAACCTTCCG
AAGGCCCATACCTGCTTTAACCGGATCGACATTCCACCATATGAGTCCTATGAGAAGCTC
TACGAGAAGCTGCTGACAGCCGTGGAGGAGACCTGCGGGTTTGCTGTGGAGTGA
40 (SEQ ID NO: 27)

Human Smurf2 nucleic acid sequence (uniprot.org/uniprot/Q9HAU4).

ATGTCTAACCCCGGACGCCGAGGAACGGGCCCGTCAAGCTGCGCCTGACAGTACTCTGT
GCAAAAAACCTGGTGAAAAAGGATTTTTTCCGACTTCTGATCCATTGCTAAGGTGGTG
45 GTTGATGGATCTGGGCAATGCCATTCTACAGATACTGTGAAGAATACGCTTGATCCAAAG
TGAATCAGCATTATGACCTGTATATTGGAAAGTCTGATTACGTTACGATCAGTGTATGG
AATCACAAGAAGATCCATAAGAAACAAGGTGCTGGATTCTCGGTTGTGTTCTGCTTCTT
TCCAATGCCATCAACCGCTCAAAAGACACTGGTTATCAGAGGTTGGATTTATGCAAACCTC
GGGCCAAATGACAATGATACAGTTAGAGGACAGATAGTAGTAAGTCTTCAGTCCAGAGAC
50 CGAATAGGCACAGGAGGACAAGTTGTGGACTGCAGTCGTTTATTTGATAACGATTTACCA
GACGGCTGGGAAGAAAGGAGAACCGCTCTGGAAGAATCCAGTATCTAAACCATATAACA
AGAATACGCAATGGGAGCGCCCAACACGACCGGCATCCGAATATTTAGCCCTGGCAGA
CCTCTTAGCTGCTTTGTTGATGAGAACACTCCAATTAGTGAACAAAATGGTGCAACATGT
GGACAGTCTTCAGATCCCAGGCTGGCAGAGAGGAGAGTCAGGTCACAACGACATAGAAAT
55 TACATGAGCAGAACACATTTACATACTCTCCAGACCTACCAGAAGGCTATGAACAGAGG
ACAACGCAACAAGGCCAGGTGTATTTCTTACATACACAGACTGGTGTGAGCACATGGCAT
GATCCAAGAGTGGCCAGGGATCTTAGCAACATCAATTGTGAAGAGCTTGGTCCATTGCCCT

CCTGGATGGGAGATCCGTAATACGGCAACAGGCAGAGTTTATTTTCGTTGACCATAACAAC
 AGAACAAACACAATTTACAGATCCTCGGCTGTCTGCTAACTTGCAATTTAGTTTTAAATCGG
 CAGAACCAATTGAAAGACCAACAGCAACAGCAAGTGGTATCGTTATGTCCTGATGACACA
 GAATGCCTGACAGTCCCAAGGTACAAGCGAGACCTGGTTCAGAACTAAAAATTTGCGG
 5 CAAGAACTTTCCCAACAACAGCCTCAGGCAGGTCAATGCCGCATTGAGGTTCCAGGGAA
 GAGATTTTTGAGGAATCATATCGACAGGTCAATGAAAAAGAGACCAAAAGATCTCTGGAAG
 CGATTAATGATAAAATTTCTGGGAGAAGAAGGCCTTGACTATGGAGGCGTTGCCAGGGAA
 TGGTTGTATCTCTTGTACATGAAATGTTGAATCCATACTATGGCCTCTTCCAGTATTCA
 AGAGATGATATTTATACATTGCAGATCAATCCTGATTCTGCAGTTAATCCGGAACATTTA
 10 TCCTATTTCCACTTTGTTGGACGAATAATGGGAATGGCTGTGTTTCATGGACATTATATT
 GATGGTGGTTTCACATTGCCCTTTTTATAAGCAATTGCTTGGGAAGTCAATTACCTTGGAT
 GACATGGAGTTAGTAGATCCGGATCTTCACAACAGTTTAGTGTGGATACTTGAGAATGAT
 ATTACAGGTGTTTTGGACCATACCTTCTGTGTTGAACATAATGCATATGGTGAAATTATT
 CAGCATGAACTTAAACCAAAATGGCAAAAGTATCCCTGTTAATGAAGAAAATAAAAAAGAA
 15 TATGTCAGGCTCTATGTGAAGTGGAGATTTTTACGAGGCATTGAGGCTCAATTCTTGGCT
 CTGCAGAAAGGATTTAATGAAGTAATTCACAACATCTGCTGAAGACATTTGATGAGAAG
 GAGTTAGAGCTCATTATTTGTGGACTTGGAAAGATAGATGTTAATGACTGGAAGGTAAC
 ACCCGGTTAAACACTGTACACCAGACAGCAACATTGTCAAATGGTCTGGAAGCTGTG
 GAGTTTTTTGATGAAGAGCGACGAGCAAGATTGCTTCAGTTTGTGACAGGATCCTCTCGA
 20 GTGCCTCTGCAGGGCTTCAAAGCATTGCAAGGTGCTGCAGGCCGAGACTCTTTACCATA
 CACCAGATTGATGCCTGCTAACAACCTGCCGAAAGCCACACTTGCTTCAATCGAATA
 GACATTCACCCTATGAAAGCTATGAAAGCTATATGAAAGCTGCTAACAGCCATTGAA
 GAAACATGTGGATTGCTGTGGAATGA

(SEQ ID NO: 28)

25

Human ITCH nucleic acid sequence (uniprot.org/uniprot/Q96J02).

GGAGTCGCCGCCGCCCGAGTTCGGTACCATGCATTTACGGTGGCCTTGTGGAGACAA
 CGCCTTAACCCAAGGAAGTGACTCAAACCTGTGAGAACTCCAGGTTTTCCAACCTATTGGT
 GGTATGTCTGACAGTGGATCACAACCTGGTTCAATGGGTAGCCTCACCATGAAATCACAG
 30 CTTCAGATCACTGTCTCTCAGCAAACTTAAGGAAAAATAAGAAGATTGGTTTGGACCA
 AGTCCTTACGTAGAGGTACAGTAGATGGACAGTCAAAGAAGACAGAAAAATGCAACAAC
 ACAACAGTCCCAAGTGGGAAGCAACCCCTTACAGTTATCGTTACCCCTGTGAGTAAATTA
 CATTTTCGTGTGTGGAGTACCAGACACTGAAATCTGATGTTTTGTGGGAAGTGTGCA
 TTAGATATTTATGAAACATTAAAGTCAAACAATATGAACTTGAAGAAGTAGTTGTGACT
 35 TTGCAGCTTGGAGGTGACAAAGAGCCAAAGAGACAATAGGAGACTTGTCAATTTGTCTT
 GATGGGCTACAGTTAGAGTCTGAAGTTGTTACCAATGGTGAACTACATGTTTCAAGAAAGT
 GCTTCTCAGAAATGATGATGGCTCCAGATCCAAGGATGAAACAAGAGTGAGCACAATGGA
 TCAGATGAACCTGAAGATGCAGGAGCTGGTGAAAAATAGGAGAGTCAAGTGGGAATTAATCT
 CCATCACTCTCAAATGGTGGTTTTAAACCTTCTAGACCTCCAAGACCTTCACGACCACCA
 40 CCACCCACCCACGTAGACCAGCATCTGTCAATGGTTCACCATCTGCCACTTCTGAAAGT
 GATGGGTCTAGTACAGGCTCTCTGCCGCCGACAAATACAAATACAAATACATCTGAAGGA
 GCAACATCTGGATTAATAATTCCTCTTACTATATCTGGAGGCTCAGGCCCTAGGCCATTA
 AATCCTGTAACCTCAAGCTCCCTTGCCACCTGGTTGGGAGCAGAGAGTGGACCAGCACGGG
 CGAGTTTACTATGTAGATCATGTTGAGAAAAAGAACACATGGGATAGACCAGAACCTCTA
 45 CCTCCTGGCTGGGAACGGCGGGTTGACAACATGGGACGTATTTATTATGTTGACCATTTC
 ACAAGAACAACAACGTGGCAGAGGCCAACACTGGAATCCGTCCGGAATATGAACAATGG
 CAGCTACAGCGTAGTCAGCTTCAAGGAGCAATGCAGCAGTTTAACCAGAGATTCAATTTAT
 GGGAAATCAAGATTTATTTGCTACATCACAAGTAAAGAATTTGATCCTCTTGGTCCATTG
 CCACCTGGATGGGAGAAGAGAACAGACAGCAATGGCAGAGTATATTTTCGTCACCAACAAC
 50 ACACGAATTACACAATGGGAAGACCCAGAAAGTCAAGGTCAATTAATGAAAAGCCCTTA
 CCTGAAGGTTGGGAAATGAGATTACAGTGGATGGAATTCATATTTTGTGGACCACAAT
 AGAAGAAGTACCACCTATATAGATCCCGCACAGGAAAATCTGCCCTAGACAATGGACCT
 CAGATAGCCTATGTTCCGGACTTCAAAGCAAAGGTTTCAAGTATTTCCGGTTCTGGTGTGAG
 CAAGTGGCCATGCCACAGCACATAAAGATTACAGTGACAAGAAAAACATTGTTTGGAGAT
 55 TCCTTTCAACAGATAATGAGCTTCAGTCCCAAGATCTGCGAAGACGTTTGTGGGTGATT
 TTTCCAGGAGAGAAGGTTTAGATTATGGAGGTGTAGCAAGAGAATGGTTCTTTCTTTTG
 TCACATGAAGTGTGAACCAATGTATTGCCCTGTTTGAATATGCAGGGAAGGATAACTAC
 TGCTTGAGATAAACCCCGCTTCTTACATCAATCCAGATCACCTGAAATATTTTCGTTTT

ATTGGCAGATTTATTGCCATGGCTCTGTTCCATGGGAAATTCATAGACACGGGTTTTTCT
TTACCATTCTATAAGCGTATCTTGAACAAACCAGTTGGACTCAAGGATTTAGAATCTATT
GATCCAGAATTTTACAATTTCTCTCATCTGGGTTAAGGAAAACAATATTGAGGAATGTGAT
TTGGAAATGTACTTCTCCGTTGACAAAGAAATCTAGGTGAAATTAAGAGTCATGATCTG
5 AAACCTAATGGTGGCAATATTCTTGTAAACAGAAGAAAAATAAGAGGAATACATCAGAATG
GTAGCTGAGTGGAGGTTGTCTCGAGGTGTGAAGAACAGACACAAGCTTTCTTTGAAGGC
TTTAATGAAATTTCTCCCCAGCAATATTTGCAATACTTTGATGCAAAGGAATTAGAGGTC
CTTTTATGTGGAATGCAAGAGATTGATTGAATGACTGGCAAAGACATGCCATCTACCGT
CATTATGCAAGGACCAGCAACAAATCATGTGGTTTTTGGCAGTTTGTAAAGAAATTGAT
10 AATGAGAAGAGAATGAGACTTCTGCAGTTTGTACTGGAACCTGCCGATTGCCAGTAGGA
GGATTGTCTGATCTCATGGGGAGCAATGGACCACAGAAATCTGCATTGAAAAAGTTGGG
AAAGAAAATTTGGCTACCCAGAAGTCATACCTGTTTTAATCGCCTGGACCTGCCACCATAC
AAGAGCTATGAGCAACTGAAGGAAAAGCTGTTGTTTGGCATAGAAGAAAACAGAAGGATTT
GGACAAGAGTAACTTCTGAGAACTTGACCATGAATGGGCAAGAATTATTTGCAATGTT
15 TGTCTTCTCTGCCTGTTGCACATCTTGTAAATTTGGACAATGGCTCTTTAGAGAGTTAT
CTGAGTGTAAATTAATGTTCTCATTTAAAAAAAAAAAAAAAAAAAAA

(SEQ ID NO: 29)

Human NEDL1 nucleic acid sequence (uniprot.org/uniprot/Q76N89).

20 GCGCATCAGGCGCTGTTGTTGGAGCCGGAACACCGTGCGACTCTGACCGAACCGGCCCCC
TCCTCGCGCACACACTCGCCGAGCCGCGCGCCCCCTCCGCGCTGACAGTGGCCGTGGCC
TCCGCTCTCTCGGGGCACCCGGCAGCCAGAGCGCAGCGAGAGCGGGCGGTGCGCCAGGGTC
CCCTCCCCAGCCAGTCCCAGGCGCCCGGTGCACTATGCGGGGCACGTGCGCCCCCAGCT
CTAATCTGCGCGCTGACAGGAGCATGATCTGTGCCAGGCCAGGGCTGCCAAGGAATTGA
25 TCGCGGTACACGTGGTGGGTCAATTATGCTGTACACCTGTGTAGTGTGAAGAATCTGTAC
CAGAACAGGTTTTTAGGCCCTGGCCGCCATGGCGTCTCCTTCTAGAACTCCCAGAGCCGA
CGCCGGTGCAGGAGCCGCTCCGATACAGCTACAACCCGACCAGTTCCACAACATGGAC
CTCAGGGGCGGCCCCCAGATGGCGTCAACCATTTCCCGCTCCACCAGCGACACTGACCTG
GTCACCTCGGACAGCCGCTCCACGCTCATGGTCAGCAGCTCCTACTATTCCATCGGGCAC
30 TCTCAGGACCTGGTCACTCCACTGGGACATAAAGGAGGAAGTGGACGCTGGGGACTGGATT
GGCATGTACCTCATTGATGAGGTCTTGTCCGAAAACCTTCTGGACTATAAAAACCGTGGA
GTCAATGGTTTCTCATCGGGGCCAGATCATCTGGAAGATCGATGCCAGCTCGTACTTTGTG
GAACCTGAACTAAGATCTGCTTCAAATACTACCATGGAGTGAGTGGGGCCCTGCGAGCA
ACCACCCCCAGTGTACGGTCAAAAACCTCGGCAGCTCCTATTTTTTAAAGCATTGGTGCT
35 GATGAGACCGTCCAAGGACAAGGAAGTCGGAGGCTGATCAGCTTCTCTCTCAGATTTCC
CAAGCCATGGGGTTGAAGAAAGGGATGTTTTTCAACCCAGACCCTTATCTGAAGATTTCC
ATTCAGCCTGGGAAACACAGCATCTTCCCCGCCCTCCCTCACCATGGACAGGAGAGGAGA
TCCAAGATCATAGGCAACACCGTGAACCCCATCTGGCAGGCCGAGCAATTCAGTTTTGTG
TCCTTGCCCACTGACGTGCTGGAAATTTAGAGTGAAGGACAAGTTTGCCAAGAGCCGCCCC
40 ATCATCAAGCGCTTCTTGGGAAAGCTGTGATGCCCGTTCAAAGACTCCTGGAGAGACAC
GCCATAGGGGATAGGGTGTGTCAGCTACACACTTGGCCGCAAGGCTTCAAACAGATCATGTG
AGTGGACAGCTGCAATTTCCGATTTGAGATCACTTCCTCCATCCACCCAGATGATGAGGAG
ATTTCCCTGAGTACCGAGCCTGAGTCAGCCCAAATTCAGGACAGCCCATGAACAACCTG
ATGGAAGCGGCAGTGGGGAACCTCGGTCTGAGGCACCAGAGTCCCTCTGAGAGCTGGAAG
45 CCAGAGCAGCTGGGTGAGGSCAGTGTCCCGATGGTCCAGGGAACCAAAGCATAGAGCTT
TCCAGACCAGCTGAGGAAGCAGCAGTCATCACGGAGGCAGGAGACCAGGGCATGCTCTCT
GTGGGACCTGAAGGGGCTGGGGAGCTCCTGGCCAGGTGCAAAAGGACATCCAGCCTGCC
CCCAGTGCAGAAGAGCTGGCCGAGCAGCTGGACCTGGGTGAGGAGGCATCAGCACTGCTG
CTGGAAGACGCTGAAGCCCCAGCCAGCACCAAGGAGGAGCCCTTGGAGGAGGAAGCAACG
50 ACCCAGAGCCGGGCTGGAAGGGAAGAAGAGGAGAAGGAGCAGGAGGAGGAGGAGATGTG
TCTACCCTGGAGCAGGGAGAGGGCAGGCTGCAGCTGCGGGCCTCGGTGAAGAGAAAAAGC
AGGCCCTGCTCCTTGCTGTGTCCGAGCTGGAGACGGTGATCGCGTCAGCCTGCGGGGAC
CCCGAGACCCCGCGGACACACTACATCCGCATCCACACCCTGCTGCACAGCATGCCCTCC
GCCCAGGGCGGCAGCGCGGCAGAGGAGGAGGACGGCGCGGAGGAGGAGTCCACCTCAAG
55 GACTCCTCGGAGAAGGATGGGCTCAGCGAGGTGGACAGGTGGCCGCTGACCCGTCTGCC
CTGGAAGAGGACAGAGAAGACGCCCGAGGGGCTACTCCAGGCACGCGCACCCCTGGCCAC
TCCGGGGGCCACTTCCCCAGCCTGGCCAAATGGCGCGGCCAGGATGGCGACACGCACCCC
AGCACCGGGAGCGAGAGCGACTCCAGCCCCAGGCAAGGCGGGGACCACAGTTGCGAGGGC

TGTGACGCGTCTGCTGCAGCCCTCGTGCTACAGCTCCTCGTGCTACAGCACGTCTGC
TACAGCAGCTCGTGCTACAGCGCTCGTGCTACAGCCCTCCTGCTACAACGGCAACAGG
TTCGCCAGCCACACGCGCTTCTCCTCCGTGGACAGCGCCAAGATCTCCGAGAGCAGGGTC
TTCTCCTCGCAAGACGACGAGGAGGAGGAGAACAGCGCGTTCGAGTCGGTACCCGACTCC
5 ATGCAGAGCCCTGAGCTGACCCCGAGTCCACGAACGGCGCTGGGCCGTGGCAAGACGAG
CTGGCCGCCCCCTAGCGGGCACGTGGAAGAAGCCCGAAGGTCTGGAATCCCCCGTGGCA
GGTCCAAGCAATCGGAGAGAAGACTGGGAAGCTCGAATTGACAGCCACGGGCGGGTCTTT
TATGTGGACCACGTGAACCGCACAAACCACCTGGCAGCGTCCGACGGCAGCAGCCACCCCG
GATGGCATGCGGAGATCGGGGTCCATCCAGCAGATGGAGCAACTCAACAGGCGGTATCAA
10 AACATTTCAGCGAACATTGCAACAGAGAGCTCCGAAGAAGATTCTGGCAGCCAAAGCTGC
GAGCAAGCCCCAGCAGGAGGAGGGCGGAGGTGGAGGGAGTGACTCAGAAGCCGAATCTTCC
CAGTCCAGCTTAGATCTAAGGAGAGAGGGGTCACTTTCTCCAGTGAACCTACAAAAAATC
ACCTTGCTGCTGCAGTCCCCAGCGGTCAAGTTTATCACCAACCCCGAGTTCTTCACTGTG
CTACACGCCAATTATAGTGCCTACCGAGTCTTACCAGTAGCACCTGCTTAAAGCACATG
15 ATTCTGAAAGTCCGACGGGATGCTCGCAATTTTGAACGCTACCAGCACAAACGGGACTTG
GTGAATTTTCATCAACATGTTCCGAGACACTCGGCTGGAAGTCCCCGGGGCTGGGAGATC
AAAACGGACCAGCAGGGAAGTCTTTTTTCGTGGACCACAACAGTTCGAGCTACCACTTTC
ATTGACCCCCGAATCCCTCTTCAGAACGGTCTGTTCTTCCCAATCATCTAACTACCGACAG
CACCTCCAGAGGCTCCGAAGTTACAGCGCCGAGAGGCTCAGAAGTTTCTAGAAACAGA
20 GGAGCCTCTTTACTGGCCAGGCCAGGACACAGCTTAGTAGCTGCTATTGGAAGCCAACAT
CAACATGAGTCATTGCCACTGGCATATAATGACAAGATTGTGGCATTTCTTCGCCAGCCA
AACATTTTTGAAATGCTGCAAGAGCGTCAGCCAAGCTTAGCAAGAAACACACACTCAGG
GAGAAAATCCATTACATTCGACTGAGGGTAATCACGGGCTTGAGAAGTTGTCCTGTGAT
GCGGATCTGGTCATTTTGTGAGTCTCTTTGAAGAAGAGATTATGTCTACGTCCCCCTG
25 CAGGCTGCCTTCCACCCTGGGTATAGCTTCTCTCCCGATGTTACCCCTGTTCTTACCT
CAGAAGTCCCAGGTTTACAGAGAGCCAGTGCAAGAGCCCCCTCCCCCTACCGAAGAGAC
TTTGAGGCCAAGCTCCGCAATTTCTACAGAAAAGTGAAGCCAAAGGATTTGGTCAGGGT
CCGGGGAAAATTAAGCTCATTATTCGCCGGGATCATTGTTGGAGGGAACCTTCAATCAG
GTGATGGCCTATTTCGCGGAAAGAGCTCCAGCGAAACAAGCTCTACGTACCTTTGTTGGA
30 GAGGAGGGCTTGGACTACAGTGGCCCCCTCGCGGGAGTTCTTCTTCTTCTGTCTCAGGAG
CTCTTCAACCCCTTACTATGAGTCTTTGAGTACTCGGCAAATGATACTTACACGCTGCAG
ATCAGCCCCATGTCCGCATTTGTAGAAAACCATCTTGAGTGGTTTCAGGTTTAGCGGTGCG
ATCCTGGGTCTGGCTCTGATCCATCAGTACCTTCTTGACGCTTTCTTACAGAGGCCCTTC
TACAAGGCACCTCTGAGACTGCCCTGTGATTTGAGTGACCTGGAATATTTGGATGAGGAA
35 TTCCACCAGATTTTGCAGTGGATGAAGGACAACAACATCACAGACATCTTAGACCTCACT
TTCAGTGTTAATGAAGAGGTTTTTGGACAGGTACCGGAAAGGAGTTGAAGTCTGGAGGA
GCCAACACACAGGTGACGGAGAAAAACAAGAAGGAGTACATCGAGCGCATGGTGAAGTGG
CGGGTGGAGCGGGCGTGGTACAGCAGACCGAGGCGTGGTGGCGGCTTCTACAGGTT
GTAGACTCGAGGCTGGTGTCCTGTGTTGATGCCAGGGAGCTGGAGCTGGTGATAGCTGGC
40 ACCGCGGAAATCGACCTAAATGACTGGCGGAATAACACTGAGTACCGGGGAGGTTACCAC
GATGGGCATCTTGTGATCCGCTGGTTCTGGGCTGCGGTGGAGCGCTTCAATAATGAGCAG
AGGCTGAGATTACTGCAGTTTGTACGGGAACATCCAGCGTGCCCTACGAAGGCTTCGCA
GCCCTCCGTGGGAGCAATGGGCTTCGGCGCTTCTGCATAGAGAAATGGGGGAAAATTAAT
TCTCTCCCCAGGGCACACATGCTTCAACCGACTGGATCTTCCACCGTATCCCTCGTAC
45 TCCATGTTGTATGAAAAGCTGTTAACAGCAGTAGAGGAAACAGCACCTTTGGACTTGAG
TGAGGACATGGAACCTCGCTGACATTTTCTGGCCAGTGACATCACCTTCTGGGATG
ATCCCCCTTTCCCTTTCCCTTAATCAACTCTCCTTTGATTTTGGTATTCCATGATTTTA
TTTTCAAC

(SEQ ID NO: 30)

50

Human NEDL2 nucleic acid sequence (uniprot.org/uniprot/ Q9P2P5).

AGAGTTCCATCAGAGCCTGCAGTGGATGAAAGACAATGATATCCATGACATCCTAGACCT
CACGTTCACTGTAACGAAGAAGTATTTGGGCAGATAACTGAACGAGAATTAAAGCCAGG
GGGTGCCAATATCCAGTTACAGAGAAGAACAAGAAGGAGTACATCGAGAGGATGGTGAA
55 GTGGAGGATTGAGAGGGGTGTTGTACAGCAAACAGAGAGCTTAGTGCGTGGCTTCTATGA
GGTGGGATGCCAGGCTGGTATCTGTTTTGATGCAAGAGAAGTGAAGATTGGTCATCGC
AGGCACAGCTGAAATAGACCTAAGTATTGGAGAAACAACACAGAATATAGAGGAGGATA
CCATGACAATCATATTGTAATTCGGTGGTCTGGGCTGCAGTGGAAAGATTCAACAATGA

ACAACGACTAAGGTTGTTACAGTTTGTACAGGCACATCCAGCATTCCCTATGAAGGATT
 TGCTTCACTCCGAGGGAGTAACGGCCCAAGAAGATTCTGTGTGGAGAAATGGGGGAAAAAT
 CACTGCTCTTCCAGAGCGCATACATGTTTAAACCGTCTGGATCTGCCTCCCTACCCATC
 CTTTTCCATGCTTTATGAAAACTGTTGACAGCAGTTGAAGAAACCAGTACTTTTGGACT
 5 TGAGTGACCTGGAAGCTGAATGCCCATCTCTGTGGACAGGCAGTTTCAGAAGCTGCCTTC
 TAGAAGAATGATTGAACATTGGAAGTTTCAAGAGGATGCTTCCTTTAGGATAAAGCTACG
 TGCTGTGTGTTTTCCAGGAACAAGTGCTCTGTACATTGGGGACTGGAGATGAGTCCTCT
 TGGAAAGGATTGGGTGAGCTTGATGCCAGGGAACAACCAACCGTCTTTCAATCAACAG
 TTCTTGACTGCCAAACTTTTTCCATTTGTTATGTTCCAAGACAAAGATGAACCCATACAT
 10 GATCAGCTCCACGGTAATTTTTAGGGACTCAGGAGAATCTTGAACCTTACCCTTGAACGT
 GGTTCAGCCAAACTGGCAGCATTGGCCCAATCTCCAAATTAGAGCAAGTTAAATAATA
 TAATAAAAGTAAATATATTTCTGAAAGTACATTCAATTAAGCCCTAAGTTATAACAGAA
 TATTCATTTCTTGCTTATGAGTGCCTGCATGGTGTGCACCATAGGTTTCCGCTTTCATGG
 GACATGAGTGAAAATGAAACCAAGTCAATATGAGGTACCTTTACAGATTGCAATAAGAT
 15 GGTCTGTGACAATGTATATGCAAGTGGTATGTGTGTAATTATGGCTAAAGACAAACCATT
 ATTCAGTGAATTACTAATGACAGATTTTATGCTTTATAATGCATGAAAACAATTTAAAA
 TAACTAGCAATTAATCACAGCATATCAGGAAAAAGTACACAGTGAGTCTGTGTTATTTTT
 TGAGGCTCATTATGTTTTATGTTCTTTAAGATGTATATAAGAACCCTACTTATCATGCTGT
 ATGTATCACTCATTCCATTTTCATGTTCCATGCATACTCGGGCATCATGCTAATATGTAT
 20 CCTTTTAAGCACTCTCAAGGAAACAAAAGGGCCTTTTATTTTTATAAAGGTAAAAAAAT
 TCCCCAAATATTTTGCACTGAATGTACCAAAGGTGAAGGGACATTACAATATGACTAACA
 GCAACTCCATCACTTGAGAAGTATAATAGAAAATAGCTTCTAAATCAAACCTTCCTTCACA
 GTGCCGTGTCTACCACTACAAGGACTGTGCATCTAAGTAATAATTTTTTAAGATTCATA
 TATGTGATAGTATGATATGCATTTATTTAAAAATGCATTAGACTCTCTCCATCCATCAAA
 25 TACTTTACAGGATGGCATTTAATACAGATATTTCTGATTTCCCCCACTGCTTTTTATTG
 TACAGCATCATTAACACTAAGCTCAGTTAAGGAGCCATCAGCAACACTGAAGAGATCAG
 TAGTAAGAATTCCATTTTCCCTCATCAGTGAAGACACCACAAATTGAAACTCAGAACTAT
 ATTTCTAAGCCTGCATTTTCACTGATGCATAATTTTCTTATTAATATTAAGAGACAGTTT
 TTCTATGGCATCTCCAAACTGCATGACATCACTAGTCTTACTTCTGCTTAATTTTATGA
 30 GAAGGTATTCTTCAATTTTAAATTGCTTTTGGGATTACTCCACATCTTGTTTATTTCTTGA
 CTAATCAGATTTTCAATAGAGTGAAGTTAAATTGGGGGTCATAAAAGCATTGGATTGACA
 TATGGTTTTGCCAGCCTATGGGTTTACAGGCATTGCCCAAACATTTCTTTGAGATCTATAT
 TTATAAGCAGCCATGGAATTCCTATTATGGGATGTTGGCAATCTTACATTTTATAGAGGT
 CATATGCATAGTTTTCATAGGTGTTTTGTAAGAACTGATTGCTCTCCTGTGAGTTAAGCT
 35 ATGTTTACTACTGGGACCCTCAAGAGGAATACCCTTATGTTACACTCCTGCCTAAAGG
 CACGTACTGCAGTGTGAAGAAATGTTCTGAAAAAGGGTTATAGAAATCTGGAATAAGAA
 AGGAAGAGCTCTCTGTATTCTATAATTGGAAGAGAAAAAAGAAAACTTTTAACTGGAA
 ATGTTAGTTTTTACTTATTGATCATGAATACAAGTATATATTTAATTTTGCAAAAAA
 AAAAAAAAAAAAAAG

40 (SEQ ID NO: 31)

In certain embodiments, the nucleic acids may encode cargo proteins having two WW domains or WW domain variants from the human ITCH protein having the nucleic acid sequence:

45 CCCTTGCCACCTGGTTGGGAGCAGAGAGTGGACCAGCACGGGCGAGTTTACTAT
 GTAGATCATGTTGAGAAAAAGAACAACATGGGATAGACCAGAACCTCTACCTCCT
 GGCTGGGAACGGCGGGTTGACAACATGGGACGTATTTATTATGTTGACCATTTC
 CAAGAACAACAACGTGGCAGAGGCCAACACTG (SEQ ID NO: 32). In other
 embodiments, the nucleic acids may encode cargo proteins having four WW domains or WW
 50 domain variants from the human ITCH protein having the nucleic acid sequence:

CCCTTGCCACCTGGTTGGGAGCAGAGAGTGGACCAGCACGGGCGAGTTTACTAT
 GTAGATCATGTTGAGAAAAGAACAACATGGGATAGACCAGAACCTCTACCTCCT
 GGCTGGGAACGGCGGGTTGACAACATGGGACGTATTTATTATGTTGACCATTTC
 CAAGAACAACAACGTGGCAGAGGCCAACACTGGAATCCGTCCGGAACCTATGAAC
 5 AATGGCAGCTACAGCGTAGTCAGCTTCAAGGAGCAATGCAGCAGTTTAACCAGA
 GATTCATTTATGGGAATCAAGATTTATTTGCTACATCACAAAGTAAAGAATTTGA
 TCCTCTTGGTCCATTGCCACCTGGATGGGAGAAGAGAACAGACAGCAATGGCAG
 AGTATATTTTCGTCAACCACAACACACGAATTACACAATGGGAAGACCCCAAG
 TCAAGGTCAATTAATGAAAAGCCCTTACCTGAAGGTTGGGAAATGAGATTCAC
 10 AGTGGATGGAATTCCATATTTTGTGGACCACAATAGAAGAACTACCACCTATATA
 GATCCCCGCACA (SEQ ID NO: 33). The nucleic acid constructs that encode the cargo
 proteins, described herein, that are fused to at least one WW domain or WW domain variant
 are non-naturally occurring, that is, they do not exist in nature.

In some embodiments the expression constructs comprise a nucleic acid sequence
 15 encoding a WW domain, or variant thereof from the nucleic acid sequence (SEQ ID NO: 23);
 (SEQ ID NO: 24); (SEQ ID NO: 25); (SEQ ID NO: 26); (SEQ ID NO: 27); (SEQ ID NO:
 28); (SEQ ID NO: 29); (SEQ ID NO: 30); (SEQ ID NO: 31); (SEQ ID NO: 32) or (SEQ ID
 NO: 33). In certain embodiments, the expression constructs encode a fusion protein
 comprising a WW domain or multiple WW domains, a nuclear localization sequence (NLS),
 20 and a Cas9 protein or variant thereof. In certain embodiments, the expression constructs
 comprise the nucleic acid sequence (SEQ ID NO: 111) or (SEQ ID NO: 112). In certain
 embodiments, the expression constructs consist of the nucleic acid sequence (SEQ ID NO:
 111) or (SEQ ID NO: 112). In certain embodiments, the expression constructs consist
 essentially of the nucleic acid sequence (SEQ ID NO: 111) or (SEQ ID NO: 112).

25 The following nucleic acid sequences encode exemplary Cas9 cargo protein
 sequences that have either 2 WW domains (SEQ ID NO: 109) or 4 WW domains (SEQ ID
 NO: 110), which were cloned into the AgeI site of the pX330 plasmid (Addgene).

ATGCCCTTGCCACCTGGTTGGGAGCAGAGAGTGGACCAGCACGGGCGAGTTTAC
 30 TATGTAGATCATGTTGAGAAAAGAACAACATGGGATAGACCAGAACCTCTACCT
 CCTGGCTGGGAACGGCGGGTTGACAACATGGGACGTATTTATTATGTTGACCATT
 TCACAAGAACAACAACGTGGCAGAGGCCAACACTGACCGGTGCCACCATGGACT
 ATAAGGACCACGACGGAGACTACAAGGATCATGATATTGATTACAAAGACGATG
 ACGATAAGATGGCCCCAAAGAAGAAGCGGAAGGTTCGGTATCCACGGAGTCCCAG

CAGCCGACAAGAAGTACAGCATCGGCCTGGACATCGGCACCAACTCTGTGGGCT
GGGCCGTGATCACCGACGAGTACAAGGTGCCAGCAAGAAATTCAGGTGCTGG
GCAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTCG
ACAGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGA
5 TACACCAGACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAG
ATGGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGG
AAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGG
TGGCCTACCACGAGAAGTACCCACCATCTACCACCTGAGAAAGAACTGGTGG
ACAGCACCGACAAGGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCACATGA
10 TCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGACAACAGCG
ACGTGGACAAGCTGTTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTTCGAGG
AAAACCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGAC
TGAGCAAGAGCAGACGGCTGGAAAATCTGATCGCCAGCTGCCCCGGCGAGAAGA
AGAATGGCCTGTTTCGGAACCTGATTGCCCTGAGCCTGGGCCTGACCCCAACTT
15 CAAGAGCAACTTCGACCTGGCCGAGGATGCCAACTGCAGCTGAGCAAGGACAC
CTACGACGACGACCTGGACAACCTGCTGGCCAGATCGGCGACCAGTACGCCGA
CCTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTG
AGAGTGAACACCGAGATCACCAAGGCCCCCTGAGCGCCTCTATGATCAAGAGA
TACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAG
20 CTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCC
GGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCC
ATCCTGGAAAAGATGGACGGCACCGAGGAACCTGCTCGTGAAGCTGAACAGAGAG
GACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGCAGCATCCCCACCAGATC
CACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTACCCATTCC
25 TGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACT
ACGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGA
GCGAGGAAACCATCACCCCTGGAACCTTCGAGGAAGTGGTGGACAAGGGCGCTT
CCGCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCCAACG
AGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTATAACG
30 AGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTGA
GCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAG
TGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACT
CCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATACC
ACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACG

AGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGA
TGATCGAGGAACGGCTGAAAACCTATGCCCACCTGTTTCGACGACAAAGTGATGA
AGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAAGCTG
ATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAG
5 TCCGACGGCTTCGCCAACAGAACTTCATGCAGCTGATCCACGACGACAGCCTG
ACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTG
CACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTG
CAGACAGTGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCC
GAGAACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGACA
10 GAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGG
GCAGCCAGATCCTGAAAGAACACCCCGTGGAACACCCAGCTGCAGAACGAG
AAGCTGTACCTGTACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGAA
CTGGACATCAACCGGCTGTCCGACTACGATGTGGACCATATCGTGCCTCAGAGCT
TTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACC
15 GGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTCTGTGAAGAAGATGAAGAAC
TACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTCGACAAT
CTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATC
AAGAGACAGCTGGTGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTG
GACTCCCGGATGAACACTAAGTACGACGAGAATGACAAGCTGATCCGGGAAGTG
20 AAAGTGATCACCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGT
TTTACAAAGTGCGCGAGATCAACAACCTACCACCACGCCCACGACGCCTACCTGA
ACGCCGTCGTGGGAACCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGT
TCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGATGATCGCCAAGAGCG
AGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCAACATCATGA
25 ACTTTTTCAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCT
GATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGGGCCGGGATTT
TGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAATATCGTGAAAAAGAC
CGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCAAGAGGAACAG
CGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTACGGCGGCTT
30 CGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAGGG
CAAGTCCAAGAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGA
AAGAAGCAGCTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAAGGGCTACAA
AGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTACTCCCTGTTTCGAGCT
GGAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGAA

ACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTA
TGAGAAGCTGAAGGGCTCCCCGAGGATAATGAGCAGAAACAGCTGTTTGTGGA
ACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCCAA
GAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACAA
5 GCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTAC
CCTGACCAATCTGGGAGCCCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGAC
CGGAAGAGGTACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCAG
AGCATCACCGGCCTGTACGAGACACGGATCGACCTGTCTCAGCTGGGAGGCGAC
AAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAG (SEQ
10 ID NO: 111)

ATGCCCTTGCCACCTGGTTGGGAGCAGAGAGTGGACCAGCACGGGCGAGTTTAC
TATGTAGATCATGTTGAGAAAAGAACAACATGGGATAGACCAGAACCTCTACCT
CCTGGCTGGGAACGGCGGGTTGACAACATGGGACGTATTTATTATGTTGACCATT
15 TCACAAGAACAACAACGTGGCAGAGGCCAACACTGGAATCCGTCCGGAACATG
AACATGGCAGCTACAGCGTAGTCAGCTTCAAGGAGCAATGCAGCAGTTTAACC
AGAGATTCATTTATGGGAATCAAGATTTATTTGCTACATCACAAAGTAAAGAATT
TGATCCTCTTGGTCCATTGCCACCTGGATGGGAGAAGAGAACAGACAGCAATGG
CAGAGTATATTTTCGTCAACCACAACACACGAATTACACAATGGGAAGACCCCAG
20 AAGTCAAGGTCAATTAATGAAAAGCCCTTACCTGAAGGTTGGGAAATGAGATT
CACAGTGGATGGAATTCCATATTTTGTGGACCACAATAGAAGAACTACCACCTAT
ATAGATCCCCGCACAGGCGGAGGAACCGGTGCCACCATGGACTATAAGGACCAC
GACGGAGACTACAAGGATCATGATATTGATTACAAAGACGATGACGATAAGATG
GCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCAGCAGCCGACAA
25 GAAGTACAGCATCGGCCTGGACATCGGCACCAACTCTGTGGGCTGGGCCGTGAT
CACCGACGAGTACAAGGTGCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGA
CCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTCGACAGCGGCGA
AACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGAC
GGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGG
30 TGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGAAAGAGGATA
AGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGTGGCCTACC
ACGAGAAGTACCCACCATCTACCACCTGAGAAAGAACTGGTGGACAGCACCG
ACAAGGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCACATGATCAAGTTCCG
GGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGACAACAGCGACGTGGACAA

GCTGTTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTTCGAGGAAAACCCCATC
AACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGC
AGACGGCTGGAAAATCTGATCGCCCAGCTGCCCCGGCGAGAAGAAGAATGGCCTG
TTCGGAAACCTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTCAAGAGCAACT
5 TCGACCTGGCCGAGGATGCCAACTGCAGCTGAGCAAGGACACCTACGACGACG
ACCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGACCTGTTTCTGGC
CGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACAC
CGAGATCACCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGATACGACGAGCA
CCACCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGCTGCCTGAGAA
10 GTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCGGCTACATTGA
CGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAAA
GATGGACGGCACCGAGGAAGTCTCGTGAAGCTGAACAGAGAGGACCTGCTGCG
GAAGCAGCGGACCTTCGACAACGGCAGCATCCCCACCAGATCCACCTGGGAGA
GCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCTGAAGGACAAC
15 CGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGGCCCTC
TGGCCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACC
ATCACCCCCTGGAACCTTCGAGGAAGTGGTGGACAAGGGCGCTTCCGCCCAGAGC
TTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCCAACGAGAAGGTGCTG
CCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTATAACGAGCTGACCAAA
20 GTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTGAGCGGCGAGCAG
AAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAG
CAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATC
TCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATACCACGATCTGCTGA
AAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACGAGGACATTCTGG
25 AAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAAC
GGCTGAAAACCTATGCCCACCTGTTTCGACGACAAAAGTGATGAAGCAGCTGAAGC
GGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAAGCTGATCAACGGCATCC
GGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAGTCCGACGGCTTCG
CCAACAGAACTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGG
30 ACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTGCACGAGCACATTG
CCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAAGG
TGGTGGACGAGCTCGTGAAAGTGATGGGCGGCACAAGCCCGAGAACATCGTGA
TCGAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGACAGAAGAACAGCCGC
GAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGCCAGATCCT

GAAAGAACACCCCGTGGAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTA
CTACCTGCAGAAATGGGCGGGATATGTACGTGGACCAGGAAGTGGACATCAACCG
GCTGTCCGACTACGATGTGGACCATATCGTGCCTCAGAGCTTTCTGAAGGACGAC
TCCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACCGGGGCAAGAGCGA
5 CAACGTGCCCTCCGAAGAGGTCGTGAAGAAGATGAAGAAGTACTGGCGGCAGCT
GCTGAACGCCAAGCTGATTACCCAGAGAAAGTTCGACAATCTGACCAAGGCCGA
GAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGGT
GGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAA
CACTAAGTACGACGAGAATGACAAGCTGATCCGGGAAGTGAAAGTGATCACCCCT
10 GAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTTTACAAAGTGCGC
GAGATCAACAACCTACCACCACGCCCACGACGCCTACCTGAACGCCGTCTGTGGGA
ACCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGAC
TACAAGGTGTACGACGTGCGGAAGATGATCGCCAAGAGCGAGCAGGAAATCGGC
AAGGCTACCGCCAAGTACTTCTTCTACAGCAACATCATGAACTTTTTCAAGACCG
15 AGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGATCGAGACAAACG
GCGAAACCGGGGAGATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGA
AAGTGCTGAGCATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTGCAGACAG
GCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGAACAGCGATAAGCTGATCG
CCAGAAAGAAGGACTGGGACCCTAAGAAGTACGGCGGCTTCGACAGCCCCACCG
20 TGGCCTATTCTGTGTGCTGGTGGTGGCCAAAGTGGAAGGGCAAGTCCAAGAAAC
TGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCG
AGAAGAATCCCATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGG
ACCTGATCATCAAGCTGCCTAAGTACTCCCTGTTCGAGCTGGAAAACGGCCGGAA
GAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGAAACGAAGTGGCCCTGCC
25 CTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGC
TCCCCCGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAAGCACTAC
CTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCC
GACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACCGGGGATAAGCCC
ATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAATCTGGGA
30 GCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGACCGGAAGAGGTACACCA
GCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCATCACCGGCCTGT
ACGAGACACGGATCGACCTGTCTCAGCTGGGAGGCGACAAAAGGCCGGCGGCCA
CGAAAAAGGCCGGCCAGGCCAAAAAAGAAAAAG (SEQ ID NO: 112)

Nucleic acids encoding any of the proteins described herein may be in any number of nucleic acid “vectors” known in the art. As used herein, a “vector” may include any nucleic acid or nucleic acid-bearing particle, cell, or organism capable of being used to transfer a nucleic acid into a host cell. The term “vector” includes both viral and nonviral products and means for introducing the nucleic acid into a cell. A “vector” can be used *in vitro*, *ex vivo*, or *in vivo*. Non-viral vectors include plasmids, cosmids, artificial chromosomes (*e.g.*, bacterial artificial chromosomes or yeast artificial chromosomes) and can comprise liposomes, electrically charged lipids (cytofectins), DNA-protein complexes, and biopolymers, for example. Viral vectors include retroviruses, lentiviruses, adeno-associated virus, pox viruses, baculovirus, reoviruses, vaccinia viruses, herpes simplex viruses, Epstein-Barr viruses, and adenovirus vectors, for example. Vectors can also comprise the entire genome sequence or recombinant genome sequence of a virus. A vector can also comprise a portion of the genome that comprises the functional sequences for production of a virus capable of infecting, entering, or being introduced to a cell to deliver nucleic acid therein.

Expression of any of the fusion proteins, described herein, may be controlled by any regulatory sequence (*e.g.* a promoter sequence) known in the art. Regulatory sequences, as described herein, are nucleic acid sequences that regulate the expression of a nucleic acid sequence. A regulatory or control sequence may include sequences that are responsible for expressing a particular nucleic acid (*i.e.* a Cas9 cargo protein) or may include other sequences, such as heterologous, synthetic, or partially synthetic sequences. The sequences can be of eukaryotic, prokaryotic or viral origin that stimulate or repress transcription of a gene in a specific or non-specific manner and in an inducible or non-inducible manner. Regulatory or control regions may include origins of replication, RNA splice sites, introns, chimeric or hybrid introns, promoters, enhancers, transcriptional termination sequences, poly A sites, locus control regions, signal sequences that direct the polypeptide into the secretory pathways of the target cell, and introns. A heterologous regulatory region is not naturally associated with the expressed nucleic acid it is linked to. Included among the heterologous regulatory regions are regulatory regions from a different species, regulatory regions from a different gene, hybrid regulatory sequences, and regulatory sequences that do not occur in nature, but which are designed by one of ordinary skill in the art.

The term operably linked refers to an arrangement of sequences or regions wherein the components are configured so as to perform their usual or intended function. Thus, a regulatory or control sequence operably linked to a coding sequence is capable of affecting the expression of the coding sequence. The regulatory or control sequences need not be

contiguous with the coding sequence, so long as they function to direct the proper expression or polypeptide production. Thus, for example, intervening untranslated but transcribed sequences can be present between a promoter sequence and the coding sequence and the promoter sequence can still be considered operably linked to the coding sequence. A

- 5 promoter sequence, as described herein, is a DNA regulatory region a short distance from the 5' end of a gene that acts as the binding site for RNA polymerase. The promoter sequence may bind RNA polymerase in a cell and/or initiate transcription of a downstream (3' direction) coding sequence. The promoter sequence may be a promoter capable of initiating transcription in prokaryotes or eukaryotes. Some non-limiting examples of eukaryotic
- 10 promoters include the cytomegalovirus (CMV) promoter, the chicken β -actin (CBA) promoter, and a hybrid form of the CBA promoter (CBh).

- In certain embodiments, the Cas9 cargo protein is expressed from the pX330 plasmid (Addgene). An exemplary nucleic acid sequence of the pX330 plasmid with the 5' AgeI cloning site underlined (single underline) and the 3' EcoRI cloning site underlined (double
- 15 underlined) is shown as (SEQ ID NO: 34). Any of the nucleic acids encoding the WW domains or WW domain variants, described herein, may be cloned, in frame, with the sequence encoding Cas9 from SEQ ID NO: 34. For example, the two ITCH WW domains or the four ITCH WW domains encoded in the nucleic acid sequences (SEQ ID NO: 32), or (SEQ ID NO: 33) may be cloned into the 5' AgeI cloning site or the 3' EcoRI cloning site. It
- 20 should be appreciated that a nucleic acid encoding any of the WW domains or WW domain variants, described herein, may be cloned into the Cas9 sequence of (SEQ ID NO: 34) and the examples provided are not meant to be limiting.

25

1 gagggcctat ttcccatgat tcttcatat ttgcatatac gatacaaggc tgtagagag
 61 ataattggaa ttaatttgac tgtaaacaca aagatattag tacaaaatac gtgacgtaga
 121 aagtaataat ttcttgggta gtttgcagtt ttaaaattat gttttaaaat ggactatcat
 181 atgcttaccg taacttgaaa gtatttcgat ttcttggcct tatatatctt GTGGAAAGGA
 241 CGAAACACCg gGTCTTCgaG AAGACctggt ttagagctaG AAAtagcaag ttaaaataag
 301 gctagtcctg tatcaacttg aaaaagtggc accgagtcgg tgcTTTTTTg ttttagagct
 30 361 agaaatagca agttaaata aggctagtcc gtTTTTagcg cgtgcgccaa ttctgcagac
 421 aaatggctct agaggtaccc gttacataac ttacggtaaa tggccgcct ggctgaccgc
 481 ccaacgaccc ccgcccattg acgtcaatag taacgccaat agggactttc cattgacgtc
 541 aatgggtgga gtatttacgg taaactgcc acttggcagt acatcaagtg tatcatatgc
 601 caagtacgcc ccctattgac gtcaatgacg gtaaatggcc cgctggcat tGtgccagt
 35 661 acatgacctt atgggacttt cctacttggc agtacatcta cgtattagtc atcgctatta
 721 ccatggtcga ggtgagcccc acgttctgct tcaactctcc catctcccc cctccccac
 781 cccaatttt gtattttatt attttttaat tattttgtgc agcgatggg gcgggggggg
 841 ggggggggcg cgcgccaggc gggcggggc gggcgaggg gcggggcggg gcgagggcga
 901 gaggtgcggc ggcagccaat cagagcggc cgctccgaaa gtttctttt atggcgaggc
 40 961 ggcggcggcg gcggccctat aaaaagcgaa gcgcgcggcg ggcgggagtc gctgcgacgc
 1021 tgccttcgcc ccgtgcccg ctcgcgcgc gcctcgcgc gcccgcccc gctctgactg
 1081 accgcgttac tcccacaggt gagcgggcgg gacggccctt ctctccggg ctgtaattag

	1141	ctgagcaaga	ggtaagggtt	taagggatgg	ttggttggtg	gggtattaat	gtttaattac
	1201	ctggagcacc	tgcctgaaat	cacttttttt	caggttGGac	cgggtgccacc	ATGGACTATA
	1261	AGGACCACGA	CGGAGACTAC	AAGGATCATG	ATATTGATTA	CAAAGACGAT	GACGATAAGA
5	1321	TGGCCCCAAA	GAAGAAGCGG	AAGGTCCGTA	TCCACGGAGT	CCCAGCAGCC	GACAAGAAGT
	1381	ACAGCATCGG	CCTGGACATC	GGCACCAACT	CTGTGGGCTG	GGCCGTGATC	ACCGACGAGT
	1441	ACAAGGTGCC	CAGCAAGAAA	TTCAAGGTGC	TGGGCAACAC	CGACCGGCAC	AGCATCAAGA
	1501	AGAACCTGAT	CGGAGCCCTG	CTGTTCGACA	GCGGCGAAAC	AGCCGAGGCC	ACCCGGCTGA
	1561	AGAGAACCGC	CAGAAGAAGA	TACACCAGAC	GGAAGAACCG	GATCTGCTAT	CTGCAAGAGA
10	1621	TCTTCAGCAA	CGAGATGGCC	AAGGTGGACG	ACAGCTTCTT	CCACAGACTG	GAAGAGTCCT
	1681	TCCTGGTGGA	AGAGGATAAG	AAGCACGAGC	GGCACCCCAT	CTTCGGCAAC	ATCGTGGACG
	1741	AGGTGGCCTA	CCACGAGAAG	TACCCACCA	TCTACCACCT	GAGAAAGAAA	CTGGTGGACA
	1801	GCACCGACAA	GGCCGACCTG	CGGCTGATCT	ATCTGGCCCT	GGCCACATG	ATCGGCTCC
	1861	GGGGCCACTT	CCTGATCGAG	GGCGACCTGA	ACCCCGACAA	CAGCGACGTG	GACAAGCTGT
15	1921	TCATCCAGCT	GGTGCAGACC	TACAACCAGC	TGTTTCGAGGA	AAACCCCATC	AACGCCAGCG
	1981	GCGTGGACGC	CAAGGCCATC	CTGTCTGCCA	GACTGAGCAA	GAGCAGACGG	CTGGAAAAATC
	2041	TGATCGCCCA	GCTGCCCGGC	GAGAAGAAGA	ATGGCCTGTT	CGGAAACCTG	ATTGCCCTGA
	2101	GCCTGGGCCT	GACCCCAAC	TTCAAGAGCA	ACTTCGACCT	GGCCGAGGAT	GCCAACTGCG
	2161	AGCTGAGCAA	GGACACCTAC	GACGACGACC	TGGACAACCT	GCTGGCCAG	ATCGGCGACC
	2221	AGTACGCCGA	CCTGTTTTCTG	GCCGCCAAGA	ACCTGTCCGA	CGCCATCCTG	CTGAGCGACA
20	2281	TCCTGAGAGT	GAACACCGAG	ATCACCAAGG	CCCCCTGAG	CGCCTCTATG	ATCAAGAGAT
	2341	ACGACGAGCA	CCACCAGGAC	CTGACCCTGC	TGAAAGCTCT	CGTGCGGCAG	CAGCTGCCTG
	2401	AGAAGTACAA	AGAGATTTTC	TTCGACCAGA	GCAAGAACGG	CTACGCCGGC	TACATTGACG
	2461	GCGGAGCCAG	CCAGGAAGAG	TTCTACAAGT	TCATCAAGCC	CATCCTGGAA	AAGATGGACG
	2521	GCACCGAGGA	ACTGCTCCTG	AAGCTGAACA	GAGAGGACCT	GCTGCGGAAG	CAGCGGACCT
25	2581	TCGACAACGG	CAGCATCCCC	CACCAGATCC	ACCTGGGAGA	GCTGCACGCC	ATTCTGCGGC
	2641	GGCAGGAAGA	TTTTTACCCA	TTCTGAAGG	ACAACCGGGA	AAAGATCGAG	AAGATCCTGA
	2701	CCTTCCGCAT	CCCCTACTAC	GTGGGCCCTC	TGGCCAGGGG	AAACAGCAGA	TTCGCCCTGA
	2761	TGACCAGAAA	GAGCGAGGAA	ACCATCACCC	CCTGGAACCT	CGAGGAAGTG	GTGGACAAGG
	2821	GCGCTTCCGC	CCAGAGCTTC	ATCGAGCGGA	TGACCAACTT	CGATAAAGAT	CTGCCCAACG
30	2881	AGAAGGTGCT	GCCCAAGCAC	AGCCTGCTGT	ACGAGTACTT	CACCGTGTAT	AACGAGCTGA
	2941	CCAAAGTGAA	ATACGTGACC	GAGGGAATGA	GAAAGCCCGC	CTTCCTGAGC	GGCGAGCAGA
	3001	AAAAGGCCAT	CGTGGACCTG	CTGTTCAAGA	CCAACCGGAA	AGTGACCGTG	AAGCAGCTGA
	3061	AAGAGGACTA	CTTCAAGAAA	ATCGAGTGCT	TCGACTCCGT	GGAAATCTCC	GGCGTGGAAAG
	3121	ATCGGTTCAA	CGCCTCCCTG	GGCACATACC	ACGATCTGCT	GAAAATTATC	AAGGACAAGG
35	3181	ACTTCTGGA	CAATGAGGAA	AACGAGGACA	TTCTGGAAGA	TATCGTGATG	ACCTGCACAC
	3241	TGTTTGAGGA	CAGAGAGATG	ATCGAGGAAC	GGCTGAAAAC	CTATGCCAC	CTGTTTCGACG
	3301	ACAAAGTGAT	GAAGCAGCTG	AAGCGGCGGA	GATACACCGG	CTGGGGCAGG	CTGAGCCGGA
	3361	AGCTGATCAA	CGGCATCCGG	GACAAGCAGT	CCGGCAAGAC	AATCCTGGAT	TTCCTGAAGT
	3421	CCGACGGCTT	CGCCAACAGA	AACTTCATGC	AGCTGATCCA	CGACGACAGC	CTGACCTTTA
40	3481	AAGAGGACAT	CCAGAAAGCG	CAGGTGTCCG	GCCAGGGCGA	TAGCCTGCAC	GAGCACATTG
	3541	CCAATCTGGC	CGGCAGCCCG	GCCATTAAAG	AGGGCATCCT	GCAGACAGTG	AAGAGCGTGG
	3601	ACGAGCTCGT	GAAAGTGATG	GGCCGGCACA	AGCCCGAGAA	CATCGTGATC	GAAATGGCCA
	3661	GAGAGAACCA	GACCACCCAG	AAGGGACAGA	AGAACAGCCG	CGAGAGAATG	AAGCGGATCG
	3721	AAGAGGGCAT	CAAAGAGCTG	GGCAGCCAGA	TCCTGAAAAG	ACACCCCGTG	GAAAACACCC
45	3781	AGCTGCAGAA	CGAGAAGCTG	TACCTGTACT	ACCTGCAGAA	TGGGCGGGAT	ATGTACGTGG
	3841	ACCAGGAAC	GGACATCAAC	CGGCTGTCCG	ACTACGATGT	GGACCATATC	GTGCCTCAGA
	3901	GCTTTCTGAA	GGACGACTCC	ATCGACAACA	AGGTGCTGAC	CAGAAGCGAC	AAGAACCAGG
	3961	GCAAGAGCGA	CAACGTGCCC	TCCGAAGAGG	TCGTGAAGAA	GATGAAGAAC	TACTGGCGGC
	4021	AGCTGCTGAA	CGCCAAGCTG	ATTACCCAGA	GAAAGTTCTG	CAATCTGACC	AAGGCCGAGA
50	4081	GAGGCGGCCT	GAGCGAATCG	GATAAGGCCG	GCTTCATCAA	GAGACAGCTG	GTGGAACCCC
	4141	GGCAGATCAC	AAAGCACGTG	GCACAGATCC	TGGACTCCCG	GATGAACACT	AAGTACGACG
	4201	AGAATGACAA	GCTGATCCGG	GAAGTGAAG	TGATCACCCCT	GAAGTCCAAG	CTGGTGTCCG
	4261	ATTTCCGGAA	GGATTTCCAG	TTTTACAAAG	TGCGCGAGAT	CAACAACACT	CACCACGCCC
	4321	ACGACGCCTA	CCTGAACGCC	GTCGTGGGAA	CCGCCCTGAT	CAAAAAGTAC	CCTAAGCTGG
55	4381	AAAGCGAGTT	CGTGTACGGC	GACTACAAGG	TGTACGACGT	GCGGAAGATG	ATCGCCAAGA
	4441	GCGAGCAGGA	AATCGGCAAG	GCTACCGCCA	AGTACTTCTT	CTACAGCAAC	ATCATGAACT
	4501	TTTTCAAGAC	CGAGATTACC	CTGGCCAACG	GCGAGATCCG	GAAGCGGCCT	CTGATCGAGA
	4561	CAAACGGCGA	AACCGGGGAG	ATCGTGTGGG	ATAAGGGCCG	GGATTTTGCC	ACCGTGGGGA
	4621	AAGTGTGAG	CATGCCCAAA	GTGAATATCG	TGAAAAAGAC	CGAGGTGCAG	ACAGGCGGCT
60	4681	TCAGCAAAGA	GTCTATCCTG	CCCAAGAGGA	ACAGCGATAA	GCTGATCGCC	AGAAAGAAGG
	4741	ACTGGGACCC	TAAGAAGTAC	GGCGGCTTCG	ACAGCCCCAC	CGTGGCCTAT	TCTGTGCTGG

	4801	TGGTGGCCAA	AGTGGAAAAG	GGCAAGTCCA	AGAAACTGAA	GAGTGTGAAA	GAGCTGCTGG
	4861	GGATCACCAT	CATGGAAAGA	AGCAGCTTCG	AGAAGAATCC	CATCGACTTT	CTGGAAGCCA
	4921	AGGGCTACAA	AGAAGTGAAA	AAGGACCTGA	TCATCAAGCT	GCCTAAGTAC	TCCCTGTTCG
5	4981	AGCTGGAAAA	CGGCCGGAAG	AGAATGCTGG	CCTCTGCCGG	CGAACTGCAG	AAGGGAACCG
	5041	AACTGGCCCT	GCCCTCCAAA	TATGTGAACT	TCCTGTACCT	GGCCAGCCAC	TATGAGAAGC
	5101	TGAAGGGCTC	CCCCGAGGAT	AATGAGCAGA	AACAGCTGTT	TGTGGAACAG	CACAAGCACT
	5161	ACCTGGACGA	GATCATCGAG	CAGATCAGCG	AGTTCTCCAA	GAGAGTGATC	CTGGCCGACG
	5221	CTAATCTGGA	CAAAGTGCTG	TCCGCCTACA	ACAAGCACCG	GGATAAGCCC	ATCAGAGAGC
10	5281	AGGCCGAGAA	TATCATCCAC	CTGTTTACCC	TGACCAATCT	GGGAGCCCCCT	GCCGCCTTCA
	5341	AGTACTTTGA	CACCACCATC	GACCCGAAGA	GGTACACCAG	CACCAAAGAG	GTGCTGGACG
	5401	CCACCCTGAT	CCACCAGAGC	ATCACCGGCC	TGTACGAGAC	ACGGATCGAC	CTGTCTCAGC
	5461	TGGGAGGCGA	CAAAAAGGCCG	GCGGCCACGA	AAAAGGCCGG	CCAGGCAAAA	AAGAAAAAGT
	5521	aa <u>gaattc</u> CT	AGAGCTCGCT	GATCAGCCTC	GACTGTGCCT	TCTAGTTGCC	AGCCATCTGT
	5581	TGTTTGCCCC	TCCCCGTGTC	CTTCCTTGAC	CCTGGAAGGT	GCCACTCCCA	CTGTCTTTTC
15	5641	CTAATAAAAT	GAGGAAATTG	CATCGCATTG	TCTGAGTAGG	TGTCATTCTA	TTCTGGGGGG
	5701	TGGGGTGGGG	CAGGACAGCA	AGGGGGAGGA	TTGGGAAGAg	AATAGCAGGC	ATGCTGGGGA
	5761	gcggccgcag	gaacccttag	tgatggagtt	ggccactccc	tctctgcgcg	ctcgtctcgt
	5821	cactgaggcc	gggcgaccaa	aggtcgcccg	acgcccgggc	tttgcccggg	cggcctcagt
	5881	gagcgagcga	gcgcgcagct	gcctgcaggg	gcgcctgatg	cggatattttc	tccttacgca
20	5941	tctgtgcggt	atttcacacc	gcatacgtca	aagcaaccat	agtacgcgcc	ctgtagcggc
	6001	gcattaagcg	cggcggtgtg	ggtggttacg	cgcagcgtga	ccgctacact	tgccagcgcc
	6061	ctagcgcccg	ctcctttcgc	tttcttccct	tcctttctcg	ccacgttcgc	cggctttccc
	6121	cgtcaagctc	taaatcgggg	gctcccttta	gggttcgatg	ttagtgtttt	acggcacctc
	6181	gaccccaaaa	aacttgattt	gggtgatggt	tcacgtagtg	ggccatcgcc	ctgatagacg
25	6241	gtttttcgcc	ctttgacgtt	ggagtccaag	ttctttaata	gtggactcct	gttccaaact
	6301	ggaacaacac	tcaaccctat	ctcgggctat	tcttttgatt	tataagggat	tttgccgatt
	6361	tcggcctatt	ggttaaaaaa	tgagctgatt	taacaaaaat	ttaacgcgaa	ttttaacaaa
	6421	atattaacgt	ttacaatttt	atggtgcact	ctcagtacaa	tctgctctga	tgcgcctag
	6481	ttaagccagc	cccgaacacc	gccaaacacc	gctgacgcgc	cctgacgggc	ttgtctgctc
30	6541	ccggcctccg	cttacagaca	agctgtgacc	gtctccggga	gctgcattgt	tcagaggttt
	6601	tcaccgtcat	caccgaaacg	cgcgagacga	aagggcctcg	tgatacgcct	atTTTTatag
	6661	gttaaatgtca	tgataataat	ggtttcttag	acgtcaggtg	gcacttttgc	gggaaatgtg
	6721	cgcggaaccc	ctatTTgttt	atTTttctaa	atacattcaa	atatgtatcc	gctcatgaga
	6781	caataaccct	gataaatgct	tcaataatat	tgaaaaagga	agagtatgag	tattcaacat
35	6841	ttccgtgtcg	cccttatcc	cttttttgcc	gcatttttgc	ttcctgtttt	tgtcaccoca
	6901	gaaacgcgtg	tgaaagtaaa	agatgctgaa	gatcagttgg	gtgcacgagt	gggttacatc
	6961	gaactggatc	tcaacagcgg	taagatcctt	gagagttttc	gccccgaaga	acgttttcca
	7021	atgatgagca	cttttaaaag	tctgctatgt	ggcgcggtat	tatcccgtat	tgacgcgggg
	7081	caagagcaac	tcggtcgcgc	catacactat	tctcagaatg	acttggttga	gtactacca
40	7141	gtcacagaaa	agcatcttac	ggatggcatg	acagtaagag	aattatgcag	tgtgtccata
	7201	accatgagt	ataacactgc	ggccaactta	cttctgacaa	cgatcggag	accgaaggag
	7261	ctaaccgctt	ttttgcacaa	catgggggat	catgtaactc	gccttgatcg	ttgggaaccg
	7321	gagctgaatg	aagccatacc	aaacgacgag	cgtgacacca	cgatgcctgt	agcaatggca
	7381	acaacgttgc	gcaaactatt	aactggcgaa	ctacttactc	tagcttcccg	gcaacaatta
45	7441	atagactgga	tggaggcgga	taaagttgca	ggaccacttc	tgcgctcggc	ccttccggct
	7501	ggctggttta	ttgctgataa	atctggagcc	ggtgagcgtg	gaagccgcgg	tatcattgca
	7561	gcactggggc	cagatggtaa	gccctcccg	atcgtagtta	tctacacgac	ggggagtcag
	7621	gcaactatgg	atgaacgaaa	tagacagatc	gctgagatag	gtgcctcact	gattaagcat
	7681	tggtaaactgt	cagaccaagt	ttactcatat	atactttaga	ttgattttaa	acttcatttt
50	7741	taatttaaaa	ggatctaggt	gaagatcctt	tttgataaat	tcatagacaa	aatcccttaa
	7801	cgtgagtttt	cgttccactg	agcgtcagac	cccgtagaaa	agatcaaagg	atcttcttga
	7861	gatccttttt	ttctgcgcgt	aatctgctgc	ttgcaaacaa	aaaaaccacc	gctaccagcg
	7921	gtggttttgt	tgcgggatca	agagctacca	actctttttc	cgaaggtaac	tggcttcagc
	7981	agagcgcaga	taccaaatac	tgtccttcta	gtgtagccgt	agttaggcca	ccacttcaag
55	8041	aactctgtag	caccgcctac	atacctcgct	ctgctaattc	tgttaccagt	ggctgctgcc
	8101	agtggcgata	agtcgtgtct	taccgggttg	gactcaagac	gatagttacc	ggataaggcg
	8161	cagcggtcgg	gctgaacggg	gggttcgtgc	acacagccca	gcttggagcg	aacgacctac
	8221	accgaactga	gatacctaca	gcgtgagcta	tgagaaagcg	ccacgcttcc	cgaagggaga
	8281	aaggcggaca	ggtatccggt	aagcggcagg	gtcgggaacag	gagagcgacg	gaggagcctt
60	8341	ccagggggaa	acgcctggta	tctttatagt	cctgtcgggt	ttcgccacct	ctgacttgag
	8401	cgtcgatttt	tgtgatgctc	gtcagggggg	cggagcctat	ggaaaaacgc	cagcaacgcg

8461 gccttttttac ggttcctggc cttttgctgg ccttttgctc acatgt (SEQ ID NO: 34)

Cells producing microvesicles containing cargo proteins

A microvesicle-producing cell of the present invention may be a cell containing any
5 of the expression constructs or any of the cargo proteins described herein. For example, an
inventive microvesicle-producing cell may contain one or more recombinant expression
constructs encoding (1) a minimal ARRDC1 protein, or PSAP (SEQ ID NO: 122) or PTAP
(SEQ ID NO: 23) motif-containing variant thereof, and (2) a cargo protein fused to at least
one WW domain, or variant thereof, under the control of a heterologous promoter. In certain
10 embodiments, the expression construct in the microvesicle producing cell encodes a cargo
protein with one or more WW domains or variants thereof. In some embodiments, the
expression construct encodes a Cas9 cargo protein or variant thereof fused to one or more
WW domains or variants thereof. In some embodiments, the expression construct encodes a
Cas9 cargo protein or variant thereof fused to at least one WW domain and at least one NLS.
15 In some embodiments, the expression construct further encodes a guide RNA (gRNA). In
some embodiments, the expression construct further encodes a TSG101 protein, or a TSG101
protein variant. It should be appreciated that the ARMMs produced by such a microvesicle
producing cell typically comprise the WW domain containing cargo proteins encoded by the
expression constructs described herein.

20 Another inventive microvesicle-producing cell may contain a recombinant expression
construct encoding (1) a minimal ARRDC1 protein, or a PSAP (SEQ ID NO: 122) or PTAP
(SEQ ID NO: 23) motif-containing variant thereof, linked to (2) a Cas9 cargo protein, or
variant thereof, under the control of a heterologous promoter. Some aspects of this invention
provide a microvesicle-producing cell that comprises a recombinant expression construct
25 encoding (1) a TSG101 protein, or a UEV domain-containing variant thereof, linked to (2) a
Cas9 cargo protein or variant thereof, under the control of a heterologous promoter.

Any of the expression constructs, described herein, may be stably inserted into the
genome of the cell. In some embodiments, the expression construct is maintained in the cell,
but not inserted into the genome of the cell. In some embodiments, the expression construct is
30 in a vector, for example, a plasmid vector, a cosmid vector, a viral vector, or an artificial
chromosome. In some embodiments, the expression construct further comprises additional
sequences or elements that facilitate the maintenance and/or the replication of the expression
construct in the microvesicle-producing cell, or that improve the expression of the fusion
protein in the cell. Such additional sequences or elements may include, for example, an

origin of replication, an antibiotic resistance cassette, a polyA sequence, and/or a transcriptional isolator. Some expression constructs suitable for the generation of microvesicle producing cells according to aspects of this invention are described elsewhere herein. Methods and reagents for the generation of additional expression constructs suitable for the generation of microvesicle producing cells according to aspects of this invention will be apparent to those of skill in the art based on the present disclosure. In some embodiments, the microvesicle producing cell is a mammalian cell, for example, a mouse cell, a rat cell, a hamster cell, a rodent cell, or a nonhuman primate cell. In some embodiments, the microvesicle producing cell is a human cell.

One skilled in the art may employ conventional techniques, such as molecular or cell biology, virology, microbiology, and recombinant DNA techniques. Exemplary techniques are explained fully in the literature. For example, one may rely on the following general texts to make and use the invention: Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, Second Edition (1989) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, and Sambrook *et al.* Third Edition (2001); *DNA Cloning: A Practical Approach*, Volumes I and II (D.N. Glover ed. 1985); *Oligonucleotide Synthesis* (M.J. Gaited. 1984); *Nucleic Acid Hybridization* (B.D. Hames & S.J. Higgins eds. (1985)); *Transcription and Translation* Hames & Higgins, eds. (1984); *Animal Cell Culture* (R.I. Freshney, ed. (1986)); *Immobilized Cells And Enzymes* (IRL Press, (1986)); Gennaro *et al.* (eds.) *Remington's Pharmaceutical Sciences*, 18th edition; B. Perbal, *A Practical Guide To Molecular Cloning* (1984); F.M. Ausubel *et al.* (eds.), *Current Protocols in Molecular Biology*, John Wiley & Sons, Inc.(updates through 2001), Coligan *et al.* (eds.), *Current Protocols in Immunology*, John Wiley & Sons, Inc.(updates through 2001); W. Paul *et al.* (eds.) *Fundamental Immunology*, Raven Press; E.J. Murray *et al.* (ed.) *Methods in Molecular Biology: Gene Transfer and Expression Protocols*, The Humana Press Inc. (1991)(especially vol.7); and J.E. Celis *et al.*, *Cell Biology: A Laboratory Handbook*, Academic Press (1994).

Delivery of ARMMs containing cargo proteins

The inventive microvesicles (*e.g.*, ARMMs) containing a cargo protein, described herein, may further have a targeting moiety. The targeting moiety may be used to target the delivery of ARMMs to specific cell types, resulting in the release of the contents of the ARMM into the cytoplasm of the specific targeted cell type. A targeting moiety may selectively bind an antigen of the target cell. For example, the targeting moiety may be a membrane-bound immunoglobulin, an integrin, a receptor, a receptor ligand, an aptamer, a

small molecule, or a variant thereof. Any number of cell surface proteins may also be included in an ARMM to facilitate the binding of an ARMM to a target cell and/or to facilitate the uptake of an ARMM into a target cell. Integrins, receptor tyrosine kinases, G-protein coupled receptors, and membrane-bound immunoglobulins suitable for use with

5 embodiments of this invention will be apparent to those of skill in the art and the invention is not limited in this respect. For example, in some embodiments, the integrin is an $\alpha 1\beta 1$, $\alpha 2\beta 1$, $\alpha 4\beta 1$, $\alpha 5\beta 1$, $\alpha 6\beta 1$, $\alpha L\beta 2$, $\alpha M\beta 2$, $\alpha IIb\beta 3$, $\alpha V\beta 3$, $\alpha V\beta 5$, $\alpha V\beta 6$, or a $\alpha 6\beta 4$ integrin. In some embodiments, the receptor tyrosine kinase is a an EGF receptor (ErbB family), insulin receptor, PDGF receptor, FGF receptor, VEGF receptor, HGF receptor, Trk receptor, Eph

10 receptor, AXL receptor, LTK receptor, TIE receptor, ROR receptor, DDR receptor, RET receptor, KLG receptor, RYK receptor, or MuSK receptor. In some embodiments, the G-protein coupled receptor is a rhodopsin-like receptor, the secretin receptor, metabotropic glutamate/pheromone receptor, cyclic AMP receptor, frizzled/smoothened receptor, CXCR4 receptor, CCR5 receptor, or beta-adrenergic receptor.

15 Any number of membrane-bound immunoglobulins, known in the art, may be used as targeting moieties to target the delivery of ARMMs containing a cargo protein to any number of target cell types. In certain embodiments, the membrane-bound immunoglobulin targeting moiety binds a tumor associated or tumor specific antigen. Some non-limiting examples of tumor antigens include, CA19-9, c-met, PD-1, CTLA-4, ALK, AFP, EGFR, Estrogen

20 receptor (ER), Progesterone receptor (PR), HER2/neu, KIT, B-RAF, S100, MAGE, Thyroglobulin, MUC-1, and PSMA (Bigbee W., *et al.* "Tumor markers and immunodiagnosis.", *Cancer Medicine*. 6th ed. Hamilton, Ontario, Canada: BC Decker Inc., 2003.; Andriole G, *et al.* "Mortality results from a randomized prostate-cancer screening trial.", *New England Journal of Medicine*, 360(13):1310–1319, 2009.; Schröder FH, *et al.*

25 "Screening and prostate-cancer mortality in a randomized European study." *New England Journal of Medicine*, 360(13):1320–1328, 2009.; Buys SS, *et al.* "Effect of screening on ovarian cancer mortality: the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Randomized Controlled Trial.", *JAMA*, 305(22):2295–2303, 2011; Cramer DW *et al.* "Ovarian cancer biomarker performance in prostate, lung, colorectal, and ovarian cancer

30 screening trial specimens." *Cancer Prevention Research*, 4(3):365–374, 2011.; Roy DM, *et al.* "Candidate prognostic markers in breast cancer: focus on extracellular proteases and their inhibitors.", *Breast Cancer*. Jul 3;6:81-91, 2014.; Tykodi SS. *et al.* "PD-1 as an emerging therapeutic target in renal cell carcinoma: current evidence." *Onco Targets Ther*. Jul 25;7:1349-59, 2014.; and Weinberg RA. *The Biology of Cancer*, Garland Science, Taylor &

Francis Group LLC, New York, NY, 2007.; the entire contents of each are incorporated herein by reference).

In certain embodiments, the membrane-bound immunoglobulin targeting moiety binds to an antigen of a specific cell type. The cell type may be a stem cell, such as a pluripotent stem cell. Some non-limiting examples of antigens specific to pluripotent stem cells include Oct4 and Nanog, which were the first proteins identified as essential for both early embryo development and pluripotency maintenance in embryonic stem cells (Nichols J, *et al.* "Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4.", *Cell*. 95:379-91, 1998; the contents of which are hereby incorporated by reference). In addition to Oct4, Sox2 and Nanog, many other pluripotent stem cell markers have been identified, including Sall4, Dax1, Essrb, Tbx3, Tcl1, Rif1, Nac1 and Zfp281 (Loh Y, *et al.* "The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells.", *Nat Genet.* 38:431-40, 2006). The membrane-bound immunoglobulin targeting moiety may also bind to an antigen of a differentiated cell type. For example, the targeting moiety may bind to an antigen specific for a lung epithelial cell to direct the delivery of ARMM cargo proteins to lung epithelial cells. As a non-limiting example, a membrane-bound immunoglobulin targeting moiety may bind to the alveolar epithelial type 1 cell specific protein RTI₄₀ or HTI₅₆ to deliver cargo proteins to alveolar epithelial type 1 cells (McElroy MC *et al.* "The use of alveolar epithelial type I cell-selective markers to investigate lung injury and repair.", *European Respiratory Journal* 24:4, 664-673, 2004; the entire contents of which are hereby incorporated by reference). As another example, the targeting moiety may bind a mucin, such as muc5ac, or muc5b. It should be appreciated that the examples of antigens provided in this application are not limiting and the targeting moiety may be any moiety capable of binding any cellular antigen known in the art.

Some aspects of this invention relate to the recognition that ARMMs are taken up by target cells, and ARMM uptake results in the release of the contents of the ARMM into the cytoplasm of the target cells. In some embodiments, the fusion protein is an agent that affects a desired change in the target cell, for example, a change in cell survival, proliferation rate, a change in differentiation stage, a change in a cell identity, a change in chromatin state, a change in the transcription rate of one or more genes, a change in the transcriptional profile, or a post-transcriptional change in gene expression of the target cell. It will be understood by those of skill in the art, that the agent to be delivered will be chosen according to the desired effect in the target cell.

The genome of the target cell may be edited by a nuclease delivered to the cell via a strategy or method disclosed herein, *e.g.*, by a RNA-programmable nuclease (*e.g.*, Cas9), a TALEN, or a zinc-finger nuclease, or a plurality or combination of such nucleases. Some non-limiting aspects of this invention relate to the recognition that ARMMs can be used to
5 deliver a cargo protein fused to at least one WW domain, or variant thereof, or a Cas9 fusion protein in ARMMs to the target cell or a population of target cells, for example, by contacting the target cell with ARMMs comprising the fusion protein to be delivered. Accordingly, some aspects of this invention provide ARMMs that comprise a fusion protein, for example, a Cas9 protein, or variant thereof, fused to a WW domain, a minimal ARRDC1 protein, or
10 variant thereof, or a TSG101 protein or variant thereof.

Using any of the nucleases, described herein, or any of the nucleases known in the art, a single- or double-strand break may be introduced at a specific site within the genome of a target cell by the nuclease, resulting in a disruption of the targeted genomic sequence. In some embodiments, the targeted genomic sequence is a nucleic acid sequence within the
15 coding region of a gene. In some embodiments, the strand break introduced by the nuclease leads to a mutation within the target gene that impairs the expression of the encoded gene product. In some embodiments, a nucleic acid is co-delivered to the cell with the nuclease. In some embodiments, the nucleic acid comprises a sequence that is identical or homologous to a sequence adjacent to the nuclease target site. In some such embodiments, the strand break
20 effected by the nuclease is repaired by the cellular DNA repair machinery to introduce all or part of the co-delivered nucleic acid into the cellular DNA at the break site, resulting in a targeted insertion of the co-delivered nucleic acid, or part thereof. In some embodiments, the insertion results in the disruption or repair of a pathogenic allele. In some embodiments, the insertion is detected by a suitable assay, *e.g.*, a DNA sequencing assay, a southern blot assay,
25 or an assay for a reporter gene encoded by the co-delivered nucleic acid, *e.g.*, a fluorescent protein or resistance to an antibiotic. In some embodiments, the nucleic acid is co-delivered by association to a supercharged protein. In some embodiments, the supercharged protein is also associated to the functional effector protein, *e.g.*, the nuclease. In some embodiments, the delivery of a nuclease to a target cell results in a clinically or therapeutically beneficial
30 disruption of the function of a gene.

In some embodiments, cells from a subject are obtained and a nuclease is delivered to the cells by a system or method provided herein *ex vivo*. In some embodiments, the treated cells are selected for those cells in which a desired nuclease-mediated genomic editing event

has been effected. In some embodiments, treated cells carrying a desired genomic mutation or alteration are returned to the subject they were obtained from.

Methods for engineering, generation, and isolation of nucleases targeting specific sequences, *e.g.*, Cas9, TALE, or zinc finger nucleases, and editing cellular genomes at specific target sequences, are well known in the art (see, *e.g.*, Mani *et al.*, Biochemical and Biophysical Research Communications 335:447-457, 2005; Perez *et al.*, Nature Biotechnology 26:808-16, 2008; Kim *et al.*, Genome Research, 19:1279-88, 2009; Urnov *et al.*, Nature 435:646-51, 2005; Carroll *et al.*, Gene Therapy 15:1463-68, 2005; Lombardo *et al.*, Nature Biotechnology 25:1298-306, 2007; Kandavelou *et al.*, Biochemical and Biophysical Research Communications 388:56-61, 2009; and Hockemeyer *et al.*, Nature Biotechnology 27(9):851-59, 2009, as well as the reference recited in the respective section for each nuclease). The skilled artisan will be able to ascertain suitable methods for use in the context of the present disclosure based on the guidance provided herein.

As another example, to augment the differentiation stage of a target cell, for example, to reprogram a differentiated target cell into an embryonic stem cell-like stage, the cell is contacted, in some embodiments, with ARMMs with reprogramming factors, for example, Oct4, Sox2, c-Myc, and/or KLF4 that are fused to at least one WW domain, or variant thereof. Similarly, to affect the change in the chromatin state of a target cell, the cell is contacted, in some embodiments, with ARMMs containing a chromatin modulator, for example, a DNA methyltransferase, or a histone deacetylase fused to at least one WW domain, or variant thereof. For another example, if survival of the target cell is to be diminished, the target cell, in some embodiments, is contacted with ARMMs comprising a cytotoxic agent, for example, a cytotoxic protein fused to at least one WW domain or variant thereof. Additional agents suitable for inclusion into ARMMs and for a ARMM-mediated delivery to a target cell or target cell population will be apparent to those skilled in the art, and the invention is not limited in this respect.

In some embodiments, the ARMMs comprising a cargo fused to a WW domain, or variant thereof are provided that further include a detectable label. Such ARMMs allow for the labeling of a target cell without genetic manipulation. Detectable labels suitable for direct delivery to target cells are known in the art, and include, but are not limited to, fluorescent proteins, fluorescent dyes, membrane-bound dyes, and enzymes, for example, membrane-bound or cytosolic enzymes, catalyzing the reaction resulting in a detectable reaction product. Detectable labels suitable according to some aspects of this invention further include

membrane-bound antigens, for example, membrane-bound ligands that can be detected with commonly available antibodies or antigen binding agents.

In some embodiments, ARMMs are provided that comprise a WW domain containing protein or a fusion protein comprising a WW domain or variant thereof to be delivered to a target cell. In some embodiments, the fusion protein is or comprises a transcription factor, a transcriptional repressor, a fluorescent protein, a kinase, a phosphatase, a protease, a ligase, a chromatin modulator, or a recombinase. In some embodiments, the protein is a therapeutic protein. In some embodiments the protein is a protein that affects a change in the state or identity of a target cell. For example, in some embodiments, the protein is a reprogramming factor. Suitable transcription factors, transcriptional repressors, fluorescent proteins, kinases, phosphatases, proteases, ligases, chromatin modulators, recombinases, and reprogramming factors may be fused to one or more WW domains to facilitate their incorporation into ARMMs and their function may be tested by any methods that are known to those skilled in the art, and the invention is not limited in this respect.

Methods for isolating the ARMMs described herein are also provided. One exemplary method includes collecting the culture medium, or supernatant, of a cell culture comprising microvesicle-producing cells. In some embodiments, the cell culture comprises cells obtained from a subject, for example, cells suspected to exhibit a pathological phenotype, for example, a hyperproliferative phenotype. In some embodiments, the cell culture comprises genetically engineered cells producing ARMMs, for example, cells expressing a recombinant ARMM protein, for example, a recombinant ARRDC1 or TSG101 protein, such as a minimal ARRDC1 or TSG101 protein fused to a Cas9 protein or variant thereof. In some embodiments, the supernatant is pre-cleared of cellular debris by centrifugation, for example, by two consecutive centrifugations of increasing *G* value (*e.g.*, 500*G* and 2000*G*). In some embodiments, the method comprises passing the supernatant through a 0.2 μm filter, eliminating all large pieces of cell debris and whole cells. In some embodiments, the supernatant is subjected to ultracentrifugation, for example, at 120,000*G* for 2 hours, depending on the volume of centrifugate. The pellet obtained comprises microvesicles. In some embodiments, exosomes are depleted from the microvesicle pellet by staining and/or sorting (*e.g.*, by FACS or MACS) using an exosome marker as described herein. Isolated or enriched ARMMs can be suspended in culture media or a suitable buffer, as described herein.

Methods of microvesicle-mediated delivery of cargos

Some aspects of this invention provide a method of delivering an agent, for example, a cargo fused to a WW domain (*e.g.*, a Cas9 protein fused to a WW domain) to a target cell. The target cell can be contacted with an ARMM comprising a minimal ARRDC1 in different
5 ways. For example, a target cell may be contacted directly with an ARMM as described herein, or with an isolated ARMM from a microvesicle producing cell. The contacting can be done *in vitro* by administering the ARMM to the target cell in a culture dish, or *in vivo* by administering the ARMM to a subject. Alternatively, the target cell can be contacted with a microvesicle producing cell as described herein, for example, *in vitro* by co-culturing the
10 target cell and the microvesicle producing cell, or *in vivo* by administering a microvesicle producing cell to a subject harboring the target cell. Accordingly, the method may include contacting the target cell with a microvesicle, for example, an ARMM containing any of the cargo proteins to be delivered, as described herein. The target cell may be contacted with a microvesicle-producing cell, as described herein, or with an isolated microvesicle that has a
15 lipid bilayer, a minimal ARRDC1 protein or variant thereof, and a cargo protein.

It should be appreciated that the target cell may be of any origin. For example, the target cell may be a human cell. The target cell may be a mammalian cell. Some non-limiting examples of a mammalian cell include a mouse cell, a rat cell, hamster cell, a rodent cell, and a nonhuman primate cell. It should also be appreciated that the target cell may be of any cell
20 type. For example, the target cell may be a stem cell, which may include embryonic stem cells, induced pluripotent stem cells (iPS cells), fetal stem cells, cord blood stem cells, or adult stem cells (*i.e.*, tissue specific stem cells). In other cases, the target cell may be any differentiated cell type found in a subject. In some embodiments, the target cell is a cell *in vitro*, and the method includes administering the microvesicle to the cell *in vitro*, or co-
25 culturing the target cell with the microvesicle-producing cell *in vitro*. In some embodiments, the target cell is a cell in a subject, and the method comprises administering the microvesicle or the microvesicle-producing cell to the subject. In some embodiments, the subject is a mammalian subject, for example, a rodent, a mouse, a rat, a hamster, or a non-human primate. In some embodiments, the subject is a human subject.

30 In some embodiments, the target cell is a pathological cell. In some embodiments, the target cell is a cancer cell. In some embodiments, the microvesicle is associated with a binding agent that selectively binds an antigen on the surface of the target cell. In some embodiments, the antigen of the target cell is a cell surface antigen. In some embodiments, the binding agent is a membrane-bound immunoglobulin, an integrin, a receptor, or a receptor

ligand. Suitable surface antigens of target cells, for example of specific target cell types, *e.g.* cancer cells, are known to those of skill in the art, as are suitable binding agents that specifically bind such antigens. Methods for producing membrane-bound binding agents, for example, membrane-bound immunoglobulin, for example, membrane-bound antibodies or
5 antibody fragments that specifically bind a surface antigen expressed on the surface of cancer cells, are also known to those of skill in the art. The choice of the binding agent will depend, of course, on the identity or the type of target cell. Cell surface antigens specifically expressed on various types of cells that can be targeted by ARMMs comprising membrane-bound binding agents will be apparent to those of skill in the art. It will be appreciated that
10 the present invention is not limited in this respect.

Co-culture systems

Some aspects of this invention provide *in vitro* cell culture systems having at least two types of cells: microvesicle producing cells, and target cells that take up the microvesicles
15 produced. Accordingly, in the co-culture systems provided herein, there is a shuffling of the contents of the microvesicles (*e.g.*, ARMMs comprising minimal ARRDC1) to the target cells. Such co-culture systems allow for the expression of a gene product or multiple gene products generated by the microvesicle producing cells in the target cells without genetic manipulation of the target cells.

20 In some embodiments, a co-culture system is provided that comprises (a) a microvesicle-producing cell population having a recombinant expression construct encoding (i) a minimal ARRDC1 protein, or variant thereof, fused to a cargo (*e.g.*, an endonuclease such as a Cas9 protein or variant thereof) under the control of a heterologous promoter, and/or (ii) a TSG101 protein or variant thereof fused to a Cas9 protein variant thereof under
25 the control of a heterologous promoter, and/or (iii) a cargo protein fused to a WW domain; and (b) a target cell population. In some embodiments, the minimal ARRDC1 variant comprises a PSAP (SEQ ID NO: 122) or PTAP (SEQ ID NO: 123) motif, and/or the TSG101 variant comprises a UEV domain. In some embodiments, the expression construct further encodes a guide RNA (gRNA) which may comprise a nucleotide sequence that complements
30 a target site to mediate binding of a nuclease (*e.g.*, a Cas9 nuclease) to a target site thereby providing the sequence specificity of the nuclease:RNA complex. In some embodiments, the host cell comprises a plurality of expression constructs encoding a plurality of minimal ARRDC1:Cas9 fusion proteins and/or TSG101:Cas9 fusion proteins and/or cargo proteins fused to a WW domain.

One exemplary application of a co-culture system as provided herein is the programming or reprogramming of a target cell without genetic manipulation. For example, in some embodiments, the target cell is a differentiated cell, for example, a fibroblast cell. In some embodiments, the microvesicle producing cells are feeder cells or non-proliferating
5 cells. In some embodiments, the microvesicle producing cells produce ARMMs comprising a reprogramming factor fused to one or more WW domains, or a plurality of reprogramming factors that are fused to one or more WW domains. In some embodiments, co-culture of the differentiated target cells with the microvesicle producing cells results in the reprogramming of the differentiated target cells to an embryonic state. In some embodiments, co-culture of
10 the differentiated target cells with the microvesicle producing cells results in the programming, or trans-differentiation, of the target cells to a differentiated cell states that is different from the original cell state of the target cells.

Another exemplary application of a co-culture system, as provided herein, is the directed differentiation of embryonic stem cells. In some embodiments, the target cells are
15 undifferentiated embryonic stem cells, and the microvesicle producing cells express one or more differentiation factors fused to one or more WW domains, for example, signaling molecules or transcription factors that trigger or facilitate the differentiation of the embryonic stem cells into differentiated cells of a desired lineage, for example neuronal cells, or mesenchymal cells.

Yet another exemplary application of a co-culture system, as provided herein, is the
20 maintenance of stem cells, for example, of embryonic stem cells or of adult stem cells in an undifferentiated state. In some such embodiments, the microvesicle producing cells express signaling molecules and/or transcription factors fused to one or more WW domains that promote stem cell maintenance and/or inhibit stem cell differentiation. The microvesicle
25 producing cells may create a microenvironment for the stem cells that mimics a naturally occurring stem cell niche.

The microvesicle-producing cell of a culture system may be a cell of any type or origin that is capable of producing any of the ARMMs described herein. For example, the microvesicle-producing cell may be a mammalian cell, examples of which include but are not
30 limited to, a cell from a rodent, a mouse, a rat, a hamster, or a non-human primate. The microvesicle-producing cell may also be from a human. One non-limiting example of a microvesicle-producing cell capable of producing an ARMM is a human embryonic kidney 293T cell. The microvesicle-producing cell may be a proliferating or a non-proliferating cell. In some embodiments, the microvesicle-producing cell is a feeder cell which supports the

growth of other cells in the culture. Feeder cells may provide attachment substrates, nutrients, or other factors that are needed for the growth of cells in culture.

The target cell of the culture system can be a cell of any type or origin, which may be contacted with an ARMM from any of the microvesicle-producing cells, described herein.

- 5 For example, the target cell may be a mammalian cell, examples of which include but are not limited to, a cell from a rodent, a mouse, a rat, a hamster, or a non-human primate. The target cell may also be from a human. The target cell may be from an established cell line (*e.g.*, a 293T cell), or a primary cell cultured *ex vivo* (*e.g.*, cells obtained from a subject and grown in culture). Target cells may be hematologic cells (*e.g.*, hematopoietic stem cells, leukocytes, 10 thrombocytes or erythrocytes), or cells from solid tissues, such as liver cells, kidney cells, lung cells, heart cells bone cells, skin cells, brain cells, or any other cell found in a subject. Cells obtained from a subject can be contacted with an ARMM from a microvesicle-producing cell and subsequently re-introduced into the same or another subject. In some embodiments, the target cell is a stem cell. The stem cell may be a totipotent stem cell that 15 can differentiate into embryonic and extraembryonic cell types. The stem cell may also be a pluripotent stem cell, a multipotent stem cell, an oligopotent stem cell or a unipotent stem cell. In other embodiments, the target cell is a differentiated cell.

Method of gene editing

- 20 Some aspects of the invention provide methods for gene editing by contacting a target cell with ARMMs that contain any of the RNA-programmable fusion proteins (*i.e.*, Cas9 fusion proteins) described herein. Other aspects of the invention provide methods for gene editing by contacting a target cell with a microvesicle-producing cell comprising a recombinant expression construct encoding any of the RNA-programmable fusion proteins 25 described herein. The RNA-guided or RNA-programmable fusion protein may be delivered to a target cell by any of the systems or methods provided herein. For example, the RNA-programmable fusion protein may contain a Cas9 nuclease, or variants thereof, one or more WW domains, or variants thereof, or optionally one or more NLSs which may be delivered to a target cell by the systems or methods provided herein.

- 30 In some embodiments, the RNA-programmable nuclease includes any of the Cas9 fusion proteins described herein. Because RNA-programmable nucleases (*i.e.*, Cas9) use RNA:DNA hybridization to determine target DNA cleavage sites, these proteins are able to cleave, in principle, any sequence specified by the guide RNA. Methods of using RNA-programmable nucleases, such as Cas9, for site-specific cleavage (*e.g.*, to modify a genome)

are known in the art (see *e.g.*, Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* 339, 819-823 (2013); Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* 339, 823-826 (2013); Hwang, W.Y. *et al.* Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nature biotechnology* 31, 227-229 (2013); Jinek, M. *et al.* RNA-programmed genome editing in human cells. *eLife* 2, e00471 (2013); Dicarlo, J.E. *et al.* Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. *Nucleic acids research* (2013); Jiang, W. *et al.* RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nature biotechnology* 31, 233-239 (2013); the entire contents of each of which are incorporated herein by reference).

Some aspects of this disclosure provide fusion proteins that have an RNA-guided or RNA-programmable fusion protein (*i.e.*, a Cas9 protein, or Cas9 variant) that can bind to a gRNA, which, in turn, binds a target nucleic acid sequence; and a DNA-editing domain. Some non-limiting examples of DNA-editing domains include, but are not limited to, nucleases, nickases, recombinases or deaminases. As one example, a deaminase domain that can deaminate a nucleobase, such as, for example, cytidine is fused to an RNA-guided or RNA-programmable fusion protein. In some embodiments, the deaminase is fused to any of the Cas9 fusion proteins, described herein. The deamination of a nucleobase by a deaminase can lead to a point mutation at the respective residue, which is referred to herein as nucleic acid editing. Cargo proteins having a Cas9 protein or Cas9 variant, a DNA editing domain, and a protein capable of facilitating the incorporation of the cargo protein into an ARMM (*e.g.*, a WW domain, a minimal ARRDC1 protein, or a TSG101 protein) can thus be used for the targeted editing of nucleic acid sequences. It should be appreciated that any number of DNA editing domains (*e.g.*, nucleases, nickases, recombinases and deaminases) known in the art may be fused to an (i) RNA-guided or RNA-programmable fusion protein (*e.g.*, Cas9 or a Cas9 variant), and (ii) one or more WW domains or WW domain variants, or (iii) a minimal ARRDC1 protein, or variant thereof, or (iv) a TSG101 protein, or variant thereof. Such fusion proteins are useful for targeted editing of DNA *in vitro*, *e.g.*, for the generation of mutant cells or animals; for the introduction of targeted mutations, *e.g.*, for the correction of genetic defects in cells *ex vivo*, *e.g.*, in cells obtained from a subject that are subsequently re-introduced into the same or another subject; and for the introduction of targeted mutations, *e.g.*, the correction of genetic defects or the introduction of deactivating mutations in disease-associated genes in a subject. It should also be appreciated that any of the cargo proteins, described herein, are useful for targeted editing of DNA *in vivo*, *e.g.*, for the generation of mutant cells in a subject. Delivery of ARMMs containing any of the fusion proteins,

described herein, may be administered to a subject by any of the methods or systems, described herein.

The methods of gene editing, described herein, may result in the correction of a genetic defect, *e.g.*, in the correction of a point mutation that leads to a loss of function in a gene product. In some embodiments, the genetic defect is associated with a disease or disorder, *e.g.*, a lysosomal storage disorder or a metabolic disease, such as, for example, type I diabetes. In some embodiments, the methods provided herein are used to introduce a deactivating point mutation into a gene or allele that encodes a gene product that is associated with a disease or disorder. For example, in some embodiments, methods are provided herein that employ an RNA-guided or RNA-programmable fusion protein (*i.e.*, a Cas9 protein, or Cas9 variant) fused to a DNA editing cargo protein and at least one WW domain, or variant thereof, or a minimal ARRDC1 protein, or variant thereof, or a TSG101 protein, or variant thereof, to introduce a deactivating point mutation into an oncogene. A deactivating mutation may, in some embodiments, generate a premature stop codon in a coding sequence, which results in the expression of a truncated gene product, *e.g.*, a truncated protein lacking a function of the full-length protein.

The purpose of the methods provide herein may be used to restore the function of a dysfunctional gene via genome editing. The cargo proteins provided herein can be validated for gene editing-based human therapeutics *in vitro*, *e.g.*, by correcting a disease-associated mutation in human cell culture. It will be understood by the skilled artisan that the cargo proteins provided herein, *e.g.*, the fusion proteins comprising a Cas9 protein or Cas9 variant, a nucleic acid editing domain, and at least one WW domain or a minimal ARRDC1 protein or a TSG101 protein, can be used to correct any single point T>C or A>G mutation. For example, deamination of the mutant C back to U corrects the mutation, and in the latter case, deamination of the C that is base-paired with the mutant G, followed by a round of replication, corrects the mutation.

An exemplary disease-relevant mutation that can be corrected by the instantly provided cargo proteins *in vitro* or *in vivo* is the H1047R (A3140G) polymorphism in the PIK3CA protein. The phosphoinositide-3-kinase, catalytic alpha subunit (PIK3CA) protein acts to phosphorylate the 3-OH group of the inositol ring of phosphatidylinositol. The PIK3CA gene has been found to be mutated in many different carcinomas, and thus it is considered to be a very potent oncogene (Lee JW *et al.* "PIK3CA gene is frequently mutated in breast carcinomas and hepatocellular carcinomas.", *Oncogene*. 2005; 24(8):1477-80; the entire contents of which are hereby incorporated by reference). In fact, the A3140G mutation

is present in several NCI-60 cancer cell lines such as the HCT116, SKOV3, and T47D cell lines, which are readily available from the American Type Culture Collection (ATCC) (Ikediobi ON *et al.* “Mutation analysis of 24 known cancer genes in the NCI-60 cell line set.”, *Mol Cancer Ther.* 2006; 5(11):2606-12).

5 In some embodiments, a cell carrying a mutation to be corrected, *e.g.*, a cell carrying a point mutation resulting in a H1047R or A3140G substitution in the PIK3CA protein are contacted with an ARMM containing (i) a Cas9 protein or Cas9 variant fused to (ii) at least one WW domain or variant thereof, or a minimal ARRDC1 protein or variant thereof, or a TSG101 protein or variant thereof, (iii) a deaminase fusion protein and an appropriately
10 designed gRNA targeting the fusion protein to the respective mutation site in the encoding PIK3CA gene. Control experiments can be performed where the gRNAs are designed to target the fusion proteins to non-C residues that are within the PIK3CA gene. Genomic DNA of the treated cells can be extracted and the relevant sequence of the PIK3CA genes PCR amplified and sequenced to assess the activities of the fusion proteins in human cell culture.

15 It will be understood that the example of correcting point mutations in PIK3CA is provided for illustration purposes, and is not meant to limit the instant disclosure. The skilled artisan will understand that the instantly disclosed DNA-editing cargo proteins, described herein, can be used to correct other point mutations and mutations associated with other cancers and with diseases other than cancer.

20 The successful correction of mutations in disease-associated genes and alleles using any of the ARMMs or fusion proteins, described herein, opens up new strategies for gene correction with applications in disease therapeutics and gene study. Site-specific nucleotide modification proteins like the disclosed Cas9 variants fused to DNA-editing domains and at least one WW protein or a minimal ARRDC1 protein or a TSG101 protein also have
25 applications in “reverse” gene therapy, where certain gene functions are purposely suppressed or abolished. In these cases, site-specifically mutating Trp (TGG), Gln (CAA and CAG), or Arg (CGA) residues to premature stop codons (TAA, TAG, TGA) can be used to abolish protein function *in vitro*, *ex vivo*, or *in vivo*.

30 The instant disclosure provides methods for the treatment of a subject diagnosed with a disease associated or caused by a mutation that can be corrected by any of the DNA editing cargo proteins provided herein. For example, in some embodiments, a method is provided that comprises administering to a subject having such a disease, (*e.g.*, a cancer associated with a PIK3CA point mutation) as described above, an effective amount of ARMMs containing any of the cargo proteins, described herein, that corrects the point mutation or

introduces a deactivating mutation into the disease-associated gene. It should be appreciated that the inventive ARMMs may be used to target the delivery of any of the cargo proteins, described herein, to any target cell, described herein. In some embodiments, the disease is a neoplastic disease. In some embodiments, the disease is a metabolic disease. In some
5 embodiments, the disease is a lysosomal storage disease. Other diseases that can be treated by correcting a mutation or introducing a deactivating mutation into a disease-associated gene will be known to those of skill in the art, and the disclosure is not limited in this respect.

In some embodiments, the genome of the target cell is edited by a nuclease delivered to the target cell via a system or method disclosed herein, *e.g.*, by delivering any of the Cas9
10 fusion proteins using any of the ARMMs or ARMM producing cells described herein. In some embodiments, a single- or double-strand break is introduced at a specific site within the genome of a target cell by a Cas9 protein, resulting in a disruption of the targeted genomic sequence. In some embodiments, the targeted genomic sequence is a nucleic acid sequence within the coding region of a gene. In some embodiments, the targeted genomic sequence is
15 a nucleic acid sequence outside the coding region of a gene, for example, the targeted genomic sequence may be within the promoter region of a gene. In some embodiments, the strand break introduced by the nuclease leads to a mutation within the target gene that impairs the expression of the encoded gene product.

A nucleic acid (*e.g.*, a gRNA) may be associated with an RNA-guided protein (*e.g.*, a
20 Cas9 protein, or Cas9 variant) fused to a DNA editing domain and at least one WW domain, or variant thereof, or a minimal ARRDC1 protein, or variant thereof, or a TSG101 protein, or variant thereof. Typically, a gRNA contains a nucleotide sequence that complements a target site, which mediates binding of the protein:RNA complex to a target site and providing the sequence specificity of the protein:RNA complex. Accordingly, a nucleic acid (*e.g.*, a
25 gRNA) may be co-expressed with any of the cargo proteins, described herein, in order to confer target sequence specificity to any of the RNA-guided fusion proteins, described herein. As one non-limiting example, a Cas9 variant fused to a WW domain may be co-expressed in a cell with a gRNA such that the gRNA associates with the Cas9 fusion protein and the Cas9 fusion protein, in complex with the gRNA, is loaded into an ARMM. In some
30 embodiments, the nucleic acid has a sequence that is identical or homologous to a sequence adjacent to the nuclease target site. In some such embodiments, the strand break effected by the nuclease is repaired by the cellular DNA repair machinery to introduce all or part of the co-delivered nucleic acid into the cellular DNA at the break site, resulting in a targeted

insertion of the co-delivered nucleic acid, or part thereof. In some embodiments, the insertion results in the disruption or repair of a pathogenic allele.

In certain embodiments, a catalytically inactive Cas9 fusion protein is used to activate or repress gene expression by fusing the inactive enzyme (that retains its gRNA-binding ability) to known regulatory domains. Cas9 variants that can be used to control gene expression have been described in detail, for example, in U.S. patent application number ,USSN 14/216,655, filed March 17, 2014 (published as US 2014-0273226 A1) by Wu F. *et al.*, entitled *Crispr/cas systems for genomic modification and gene modulation*, and in PCT application number PCT/US2013/074736, filed on December 12, 2013 (published as WO 2014/093655 A2) by Zhang F. *et al.*, entitled *Engineering and optimization of systems, methods and compositions for sequence manipulation with functional domains*; the entire contents of each are incorporated herein by reference. For example, a catalytically inactive Cas9 fusion protein may be fused to a transcriptional activator (*e.g.* VP64). In certain embodiments, any of the Cas9 fusion proteins described herein may be when fused to a transcriptional activator to up-regulating gene transcription of targeted genes to enhance expression. In some embodiments, a catalytically inactive Cas9 fusion protein may be fused to a transcriptional repressor (*e.g.* KRAB). In certain embodiments, any of the Cas9 fusion proteins described herein may be fused to a transcriptional repressor to down-regulate gene transcription of targeted genes to reduce expression. In some embodiments, the delivery of a nuclease to a target cell results in a clinically or therapeutically beneficial disruption or enhancement of the function of a gene. It should be appreciated that the methods described herein are not meant to be limiting and may include any method of using Cas9 that is well known in the art.

The function and advantage of these and other embodiments of the present invention will be more fully understood from the Examples below. The following Examples are intended to illustrate the benefits of the present invention and to describe particular embodiments, but are not intended to exemplify the full scope of the invention. Accordingly, it will be understood that the Examples are not meant to limit the scope of the invention.

Examples

Example 1: Minimal ARRDC1 drives ARMMs formation and budding as efficiently as full-length ARRDC1 protein

An ARRDC1 construct was made that contains the arrestin domain, PSAP (SEQ ID NO: 122) motif and the two PPXY motifs. This “minimal” ARRDC1 is about 330 amino

acid long (100 amino acids shorter than the full-length ARRDC1 (**FIG. 1** and **FIG. 5**). When expressed in HEK293T cells, the minimal ARRDC1 buds into EVs as efficiently as the full-length ARRDC1 (**FIG. 2B**). As a negative control, another ARRDC1 construct that is of a similar size but lacks part of the N-terminal arrestin domain did not bud. Importantly, the number of extracellular vesicles made by minimal ARRDC1 expression is comparable with that of the full-length ARRDC1 (**FIG. 2C**). These data indicate that the minimal ARRDC1 is able to drive ARMMs formation and budding as efficiently as the full-length ARRDC1 protein.

10 *Example 2: Minimal ARRDC1 in packaging cargos into ARMMs*

The ability of minimal ARRDC1 in packaging cargos into ARMMs was tested. A fusion construct of minimal ARRDC1 to the Cas9 protein (**FIG. 3A**) was made. When expressed in HEK293T cells, the miniARRDC1-Cas9 protein is able to bud into the extracellular vesicles (EVs), whereas the Cas9 fusion to full length ARRDC1 did not bud out (**FIG. 3B**). Moreover, the guide RNA (gRNA) associated with Cas9 was much more enriched in ARMMs produced from miniARRDC1-Cas9 fusion protein than the control Cas9. Importantly, the miniARRDC1-Cas9 fusion protein maintains efficient gene editing activity as evidenced by an assay targeting the GFP DNA locus (**FIG. 4**). These results indicate that the minimal ARRDC1 is able to package Cas9 and associated gRNA into ARMMs via direction.

References

1. Hurley JH, Boura E, Carlson LA, & Rozycki B (2010) Membrane budding. Cell 143:875-887.
- 25 2. Thery C, Ostrowski M, & Segura E (2009) Membrane vesicles as conveyors of immune responses. Nat Rev Immunol 9:581-593.
3. Henne WM, Buchkovich NJ, & Emr SD (2011) The ESCRT pathway. Dev Cell 21:77-91.
4. Katzmann DJ, Odorizzi G, & Emr SD (2002) Receptor downregulation and multivesicular-body sorting. Nat Rev Mol Cell Biol 3:893-905.
- 30 5. Babst M, Odorizzi G, Estepa EJ, & Emr SD (2000) Mammalian tumor susceptibility gene 101 (TSG101) and the yeast homologue, Vps23p, both function in late endosomal trafficking. Traffic 1:248-258.

6. Lu Q, Hope LW, Brasch M, Reinhard C, & Cohen SN (2003) TSG101 interaction with HRS mediates endosomal trafficking and receptor down-regulation. *Proc Natl Acad Sci U S A* 100:7626-7631.
7. Pornillos O, Alam SL, Davis DR, & Sundquist WI (2002) Structure of the Tsg101 UEV domain in complex with the PTAP motif of the HIV-1 p6 protein. *Nat Struct Biol* 9:812-817.
8. Pornillos O, Alam SL, Rich RL, Myszka DG, Davis DR, & Sundquist WI (2002) Structure and functional interactions of the Tsg101 UEV domain. *EMBO J* 21:2397-2406.
9. Sundquist WI, Schubert HL, Kelly BN, Hill GC, Holton JM, & Hill CP (2004) Ubiquitin recognition by the human TSG101 protein. *Mol Cell* 13:783-789.
10. Bache KG, Brech A, Mehlum A, & Stenmark H (2003) Hrs regulates multivesicular body formation via ESCRT recruitment to endosomes. *J Cell Biol* 162:435-442.
11. Pornillos O, Higginson DS, Stray KM, Fisher RD, Garrus JE, Payne M, He GP, Wang HE, Morham SG, & Sundquist WI (2003) HIV Gag mimics the Tsg101-recruiting activity of the human Hrs protein. *J Cell Biol* 162:425-434.
12. von Schwedler UK, Stuchell M, Muller B, Ward DM, Chung HY, Morita E, Wang HE, Davis T, He GP, Cimbara DM, *et al.* (2003) The protein network of HIV budding. *Cell* 114:701-713.
13. Hurley JH & Stenmark H (2011) Molecular mechanisms of ubiquitin-dependent membrane traffic. *Annu Rev Biophys* 40:119-142.
14. Schorey JS & Bhatnagar S (2008) Exosome function: from tumor immunology to pathogen biology. *Traffic* 9:871-881.
15. Thery C, Zitvogel L, & Amigorena S (2002) Exosomes: composition, biogenesis and function. *Nat Rev Immunol* 2:569-579.
16. Bieniasz PD (2009) The cell biology of HIV-1 virion genesis. *Cell Host Microbe* 5:550-558.
17. Demirov DG & Freed EO (2004) Retrovirus budding. *Virus Res* 106:87-102.
18. Morita E & Sundquist WI (2004) Retrovirus budding. *Annu Rev Cell Dev Biol* 20:395-425.
19. Garrus JE, von Schwedler UK, Pornillos OW, Morham SG, Zavitz KH, Wang HE, Wettstein DA, Stray KM, Cote M, Rich RL, *et al.* (2001) Tsg101 and the vacuolar protein sorting pathway are essential for HIV-1 budding. *Cell* 107:55-65.

20. VerPlank L, Bouamr F, LaGrassa TJ, Agresta B, Kikonyogo A, Leis J, & Carter CA (2001) Tsg101, a homologue of ubiquitin-conjugating (E2) enzymes, binds the L domain in HIV type 1 Pr55(Gag). *Proc Natl Acad Sci U S A* 98:7724-7729.
21. Martin-Serrano J, Zang T, & Bieniasz PD (2001) HIV-1 and Ebola virus encode small
5 peptide motifs that recruit Tsg101 to sites of particle assembly to facilitate egress. *Nat Med* 7:1313-1319.
22. Martin-Serrano J, Zang T, & Bieniasz PD (2003) Role of ESCRT-I in retroviral budding. *J Virol* 77:4794-4804.
23. Demirov DG, Ono A, Orenstein JM, & Freed EO (2002) Overexpression of the N-
10 terminal domain of TSG101 inhibits HIV-1 budding by blocking late domain function. *Proc Natl Acad Sci U S A* 99:955-960.
24. Gottlinger HG, Dorfman T, Sodroski JG, & Haseltine WA (1991) Effect of mutations affecting the p6 gag protein on human immunodeficiency virus particle release. *Proc Natl Acad Sci U S A* 88:3195-3199.
25. Huang M, Orenstein JM, Martin MA, & Freed EO (1995) p6Gag is required for particle
15 production from full-length human immunodeficiency virus type 1 molecular clones expressing protease. *J Virol* 69:6810-6818.
26. Freed EO & Mouland AJ (2006) The cell biology of HIV-1 and other retroviruses. *Retrovirology* 3:77.
27. Martin-Serrano J & Neil SJ Host factors involved in retroviral budding and release. *Nat Rev Microbiol* 9:519-531.
20
28. Rauch S & Martin-Serrano J (2011) Multiple interactions between the ESCRT machinery and arrestin-related proteins: implications for PPXY-dependent budding. *J Virol* 85:3546-3556.
29. Ono A & Freed EO (2004) Cell-type-dependent targeting of human immunodeficiency
25 virus type 1 assembly to the plasma membrane and the multivesicular body. *J Virol* 78:1552-1563.
30. Pisitkun T, Shen RF, & Knepper MA (2004) Identification and proteomic profiling of exosomes in human urine. *Proc Natl Acad Sci U S A* 101:13368-13373.
31. Welton JL, Khanna S, Giles PJ, Brennan P, Brewis IA, Staffurth J, Mason MD, & Clayton A (2010) Proteomics analysis of bladder cancer exosomes. *Mol Cell Proteomics* 9:1324-1338.
30
32. Mathivanan S, Lim JW, Tauro BJ, Ji H, Moritz RL, & Simpson RJ (2009) Proteomics analysis of A33 immunoaffinity-purified exosomes released from the human colon

- tumor cell line LIM1215 reveals a tissue-specific protein signature. *Mol Cell Proteomics* 9:197-208.
33. Razi M & Futter CE (2006) Distinct roles for Tsg101 and Hrs in multivesicular body formation and inward vesiculation. *Mol Biol Cell* 17:3469-3483.
 - 5 34. Hammarstedt M & Garoff H (2004) Passive and active inclusion of host proteins in human immunodeficiency virus type 1 gag particles during budding at the plasma membrane. *J Virol* 78:5686-5697.
 35. Babst M (2005) A protein's final ESCRT. *Traffic* 6:2-9.
 36. Scott A, Chung HY, Gonciarz-Swiatek M, Hill GC, Whitby FG, Gaspar J, Holton JM,
10 Viswanathan R, Ghaffarian S, Hill CP, *et al.* (2005) Structural and mechanistic studies of VPS4 proteins. *EMBO J* 24:3658-3669.
 37. Alvarez CE (2008) On the origins of arrestin and rhodopsin. *BMC Evol Biol* 8:222.
 38. Lefkowitz RJ & Shenoy SK (2005) Transduction of receptor signals by beta-arrestins. *Science* 308:512-517.
 - 15 39. Draheim KM, Chen HB, Tao Q, Moore N, Roche M, & Lyle S (2010) ARRDC3 suppresses breast cancer progression by negatively regulating integrin beta4. *Oncogene* 29:5032-5047.
 40. Nabhan JF, Pan H, & Lu Q (2010) Arrestin domain-containing protein 3 recruits the NEDD4 E3 ligase to mediate ubiquitination of the beta2-adrenergic receptor. *EMBO Rep* 11:605-611.
20
 41. Chantry A (2011) WWP2 ubiquitin ligase and its isoforms: new biological insight and promising disease targets. *Cell Cycle* 10:2437-2439.
 42. Rotin D & Kumar S (2009) Physiological functions of the HECT family of ubiquitin ligases. *Nat Rev Mol Cell Biol* 10:398-409.
 - 25 43. Denzer K, Kleijmeer MJ, Heijnen HF, Stoorvogel W, & Geuze HJ (2000) Exosome: from internal vesicle of the multivesicular body to intercellular signaling device. *J Cell Sci* 113 Pt 19:3365-3374.
 44. Komada M & Soriano P (1999) Hrs, a FYVE finger protein localized to early endosomes, is implicated in vesicular traffic and required for ventral folding
30 morphogenesis. *Genes Dev* 13:1475-1485.
 45. Ono A, Demirov D, & Freed EO (2000) Relationship between human immunodeficiency virus type 1 Gag multimerization and membrane binding. *J Virol* 74:5142-5150.
 46. Fujii K, Hurley JH, & Freed EO (2007) Beyond Tsg101: the role of Alix in 'ESCRTing' HIV-1. *Nat Rev Microbiol* 5:912-916.

47. Wehman AM, Poggioli C, Schweinsberg P, Grant BD, & Nance J (2011) The P4-ATPase TAT-5 Inhibits the Budding of Extracellular Vesicles in *C. elegans* Embryos. *Curr Biol* 21:1951-1959.
48. Skog J, Wurdinger T, van Rijn S, Meijer DH, Gainche L, Sena-Esteves M, Curry WT, Jr., Carter BS, Krichevsky AM, & Breakefield XO (2008) Glioblastoma microvesicles transport RNA and proteins that promote tumour growth and provide diagnostic biomarkers. *Nat Cell Biol* 10:1470-1476.
49. Valadi H, Ekstrom K, Bossios A, Sjostrand M, Lee JJ, & Lotvall JO (2007) Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells. *Nat Cell Biol* 9:654-659.

All publications, patents and sequence database entries mentioned herein, including those items listed above, are hereby incorporated by reference in their entirety as if each individual publication or patent was specifically and individually indicated to be incorporated by reference. In case of conflict, the present application, including any definitions herein, will control.

Equivalents and Scope

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. The scope of the present invention is not intended to be limited to the above description, but rather is as set forth in the appended claims.

In the claims articles such as “a,” “an,” and “the” may mean one or more than one unless indicated to the contrary or otherwise evident from the context. Claims or descriptions that include “or” between one or more members of a group are considered satisfied if one, more than one, or all of the group members are present in, employed in, or otherwise relevant to a given product or process unless indicated to the contrary or otherwise evident from the context. The invention includes embodiments in which exactly one member of the group is present in, employed in, or otherwise relevant to a given product or process. The invention also includes embodiments in which more than one, or all of the group members are present in, employed in, or otherwise relevant to a given product or process.

Furthermore, it is to be understood that the invention encompasses all variations, combinations, and permutations in which one or more limitations, elements, clauses, descriptive terms, *etc.*, from one or more of the claims or from relevant portions of the

description is introduced into another claim. For example, any claim that is dependent on another claim can be modified to include one or more limitations found in any other claim that is dependent on the same base claim. Furthermore, where the claims recite a composition, it is to be understood that methods of using the composition for any of the purposes disclosed herein are included, and methods of making the composition according to any of the methods of making disclosed herein or other methods known in the art are included, unless otherwise indicated or unless it would be evident to one of ordinary skill in the art that a contradiction or inconsistency would arise.

Where elements are presented as lists, *e.g.*, in Markush group format, it is to be understood that each subgroup of the elements is also disclosed, and any element(s) can be removed from the group. It is also noted that the term “comprising” is intended to be open and permits the inclusion of additional elements or steps. It should be understood that, in general, where the invention, or aspects of the invention, is/are referred to as comprising particular elements, features, steps, *etc.*, certain embodiments of the invention or aspects of the invention consist, or consist essentially of, such elements, features, steps, *etc.* For purposes of simplicity those embodiments have not been specifically set forth *in haec verba* herein. Thus, for each embodiment of the invention that comprises one or more elements, features, steps, *etc.*, the invention also provides embodiments that consist or consist essentially of those elements, features, steps, *etc.*

Where ranges are given, endpoints are included. Furthermore, it is to be understood that unless otherwise indicated or otherwise evident from the context and/or the understanding of one of ordinary skill in the art, values that are expressed as ranges can assume any specific value within the stated ranges in different embodiments of the invention, to the tenth of the unit of the lower limit of the range, unless the context clearly dictates otherwise. It is also to be understood that unless otherwise indicated or otherwise evident from the context and/or the understanding of one of ordinary skill in the art, values expressed as ranges can assume any subrange within the given range, wherein the endpoints of the subrange are expressed to the same degree of accuracy as the tenth of the unit of the lower limit of the range.

In addition, it is to be understood that any particular embodiment of the present invention may be explicitly excluded from any one or more of the claims. Where ranges are given, any value within the range may explicitly be excluded from any one or more of the claims. Any embodiment, element, feature, application, or aspect of the compositions and/or methods of the invention, can be excluded from any one or more claims. For purposes of

brevity, all of the embodiments in which one or more elements, features, purposes, or aspects is excluded are not set forth explicitly herein.