Supporting Information

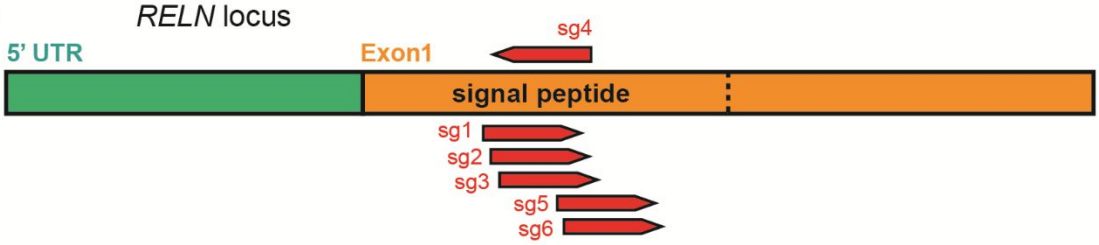
Purification of an intact human protein overexpressed from its endogenous locus via direct genome engineering

Jihyeon Yu, Eunju Cho, Yeon-Gil Choi, You Kyeong Jeong, Yongwoo Na, Jong-Seo Kim, Sung-Rae Cho, Jae-Sung Woo, Sangsu Bae

Contents

Supplementary Figure 1. Design of sgRNAs and determination of gene editing efficiency. (A) Using Cas-designer and Cas-OFFinder software (http://www.rgenome.net/), we designed six sgRNAs targeting the endogenous SP coding region of the *RELN* gene. (B) A Cas9 expression plasmid was transfected with each sgRNAs expression plasmid into HEK293E cells and the efficiency of the resulting gene editing was analyzed by targeted deep sequencing.

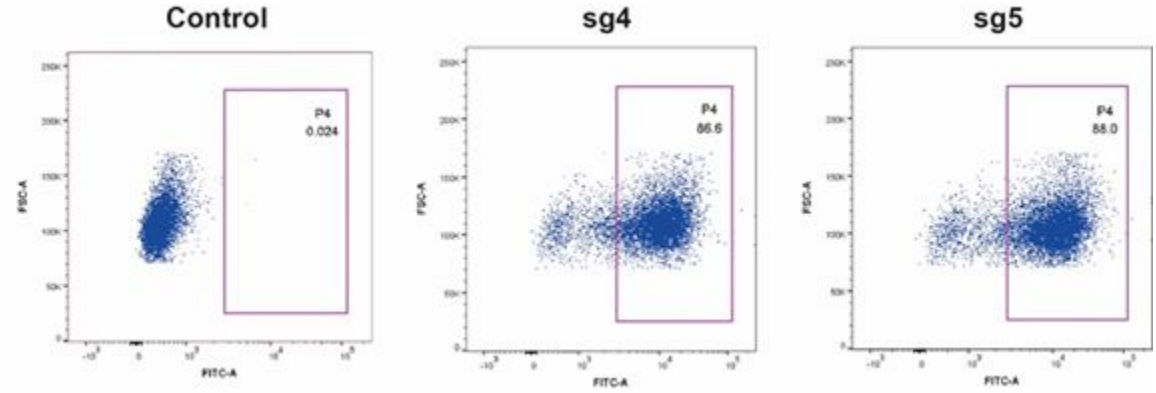
A



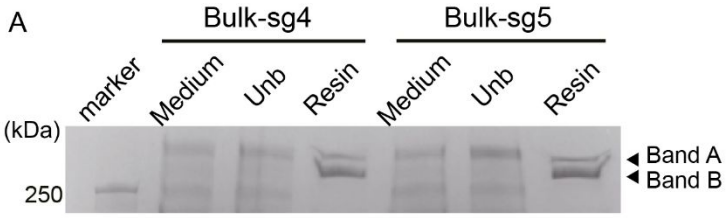
B

Name	Target sequences (5' to 3')	Direction	GC contents (%, w/o PAM)	Mismatch 0	Mismatch 1	Mismatch 2	Mutation frequency (%)
sg1	TTTCCTCCTAGCGCTGTTGCTGG	+	55	1	0	0	20.4%
sg2	TTCCTCCTAGCGCTGTTGCTGGG	+	55	1	0	0	27.0%
sg3	TCCTCCTAGCGCTGTTGCTGGGG	+	60	1	0	0	23.7%
sg4	CCCCCAGCAACAGCGCTAGGAGG	-	70	1	0	0	53.5%
sg5	TGTTGCTGGGGGCGACGCTGAGG	+	70	1	0	0	37.6%
sg6	GTTGCTGGGGGCGACGCTGAGGG	+	70	1	0	0	27.6%

Supplementary Figure 2. Knock-in cell sorting by flow cytometry. Edited cells were selected by their resistance to hygromycin, and GFP-positive cells were sorted by flow cytometry. After antibiotic selection, most cells that had been transfected with the three plasmids (Cas9 encoding, sgRNA4/5 encoding, and HDR donor) expressed GFP, in contrast to a control that had only been transfected with the HDR donor plasmid.



Supplementary Figure 3. Identification of proteins immunoprecipitated with anti-FLAG antibodies from the medium harvested from bulk cell cultures. (A) Genome-edited GFP-positive bulk cell populations produced using two different sgRNAs (sg4 and sg5) were individually cultured, and samples of the culture media were subjected to a pull-down assay using FLAG M1 antibody resin. On the SDS-PAGE gel, resin-bound fractions (in the 4th and 7th lanes) contain enriched proteins larger than 250 kDa, which are likely Reelin and/or its derived fragments. B. The two bands running above 250 kDa on the SDS-PAGE gel, which are indicated by arrowheads in A, were separately analyzed by LC-MS/MS. (B) The names of the proteins that were identified, and the number of times they were detected, are shown for the upper (Band A) and lower (Band B) bands in the left and right panels, respectively.



B

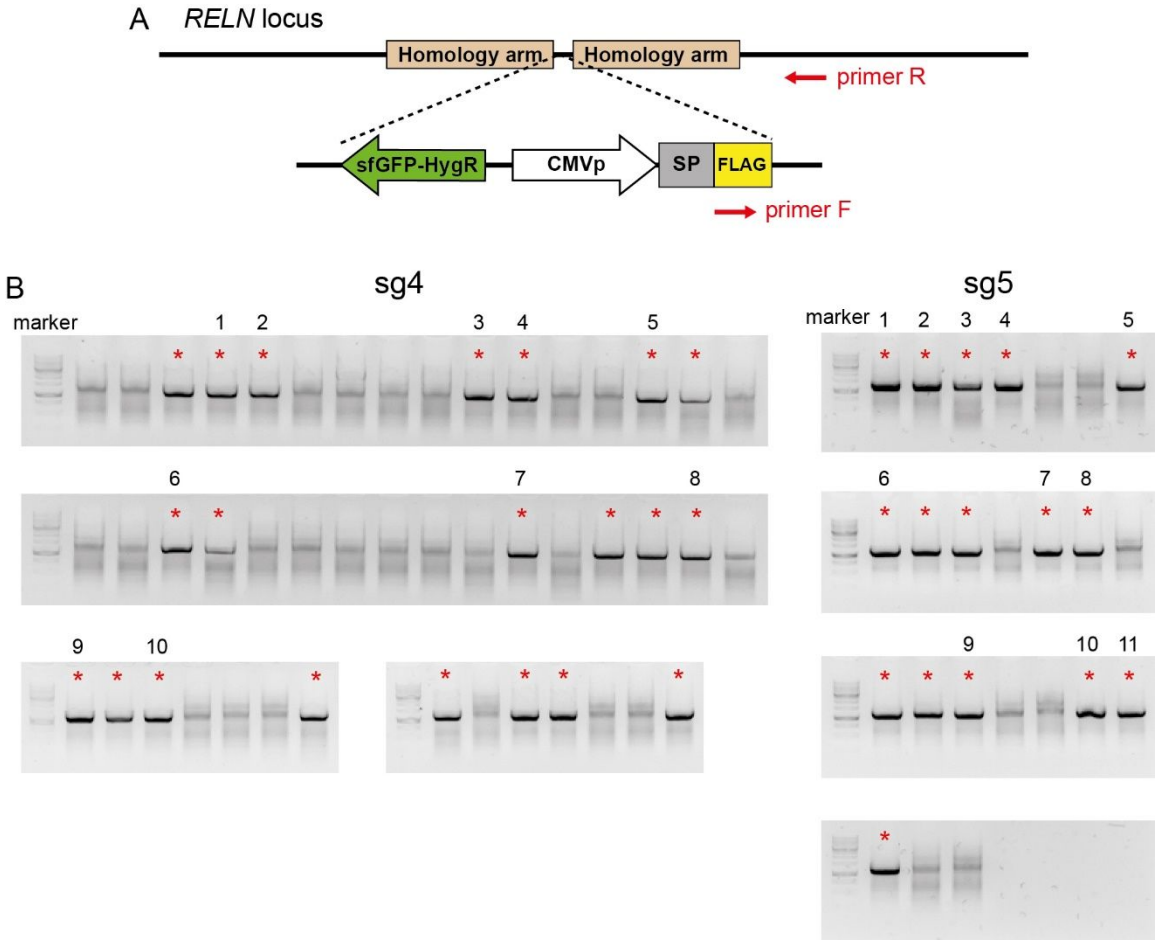
Band A

Protein	Number of peptides	Total spectral counts
RELN_HUMAN	104	359
sp|P00761|TRYP_PIG	2	15
HS90B_HUMAN	3	8
EF1A1_HUMAN	3	8
G3P_HUMAN	3	7
VIME_HUMAN	3	7
TBA1A_HUMAN	3	6
TCPH_HUMAN	5	6
EF2_HUMAN	3	6
NPM_HUMAN	1	5

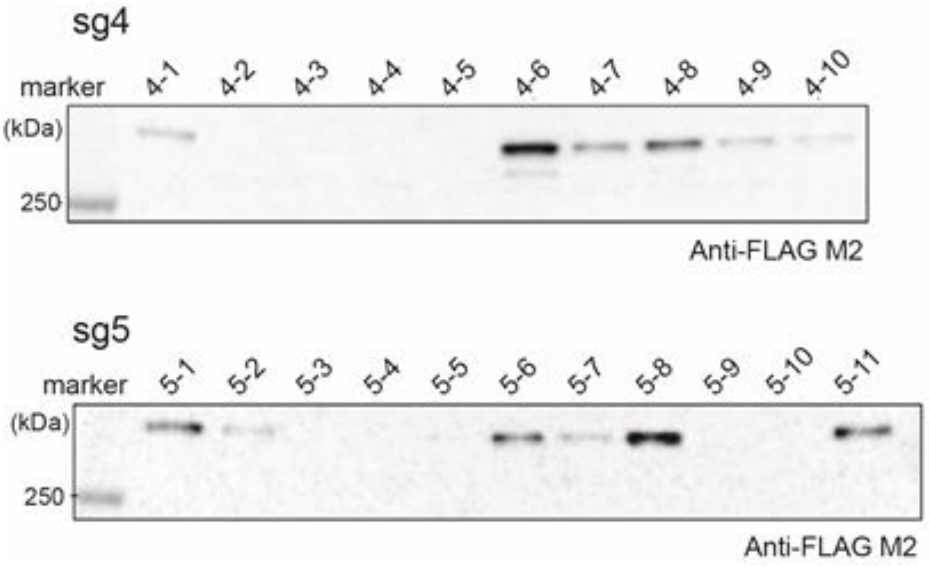
Band B

Protein	Number of peptides	Total spectral counts
RELN_HUMAN	116	580
sp|P00761|TRYP_PIG	2	16
ALBU_HUMAN	4	8
SVEP1_HUMAN	7	7
VIME_HUMAN	2	6
EF2_HUMAN	4	6
G3P_HUMAN	3	6
EF1A1_HUMAN	3	5
HNRH1_HUMAN	2	5
ACTA_HUMAN	4	5

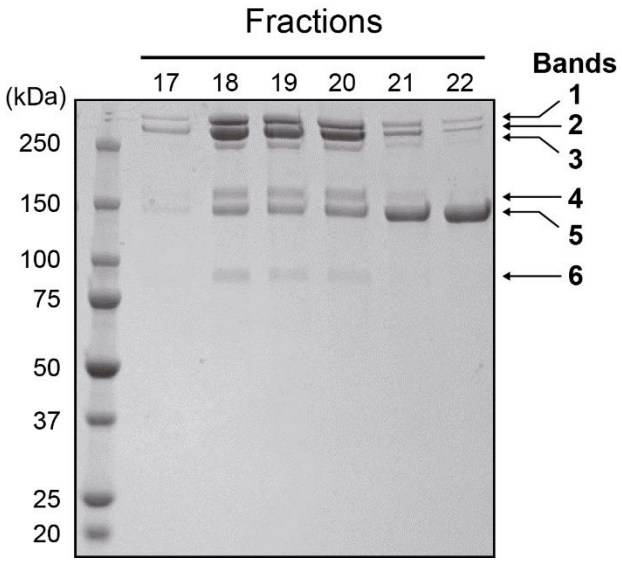
Supplementary Figure 4. Confirmation of knock-in cell lines. To confirm the presence of knock-in sequences, (A) an HDR-specific primer pair was designed and (B) PCR were performed with the primer pair that selectively amplifies DNAs between integrated HDR donor and flanking genomic sequences for selected 21 single cell-derived clonal cell lines. Used primers were 5'-GATTACAAGGATGACGATGA-3' (Primer F) and 5'- AATAGTGCCTGTCGCTGCTT-3' (Primer R).



Supplementary Figure 5. Expression of FLAG-tagged Reelin protein in single cell-derived clonal cell lines. The culture medium was harvested from 21 isolated single cell-derived clonal cell lines and Western blotting was performed using anti-FLAG M2 antibody.

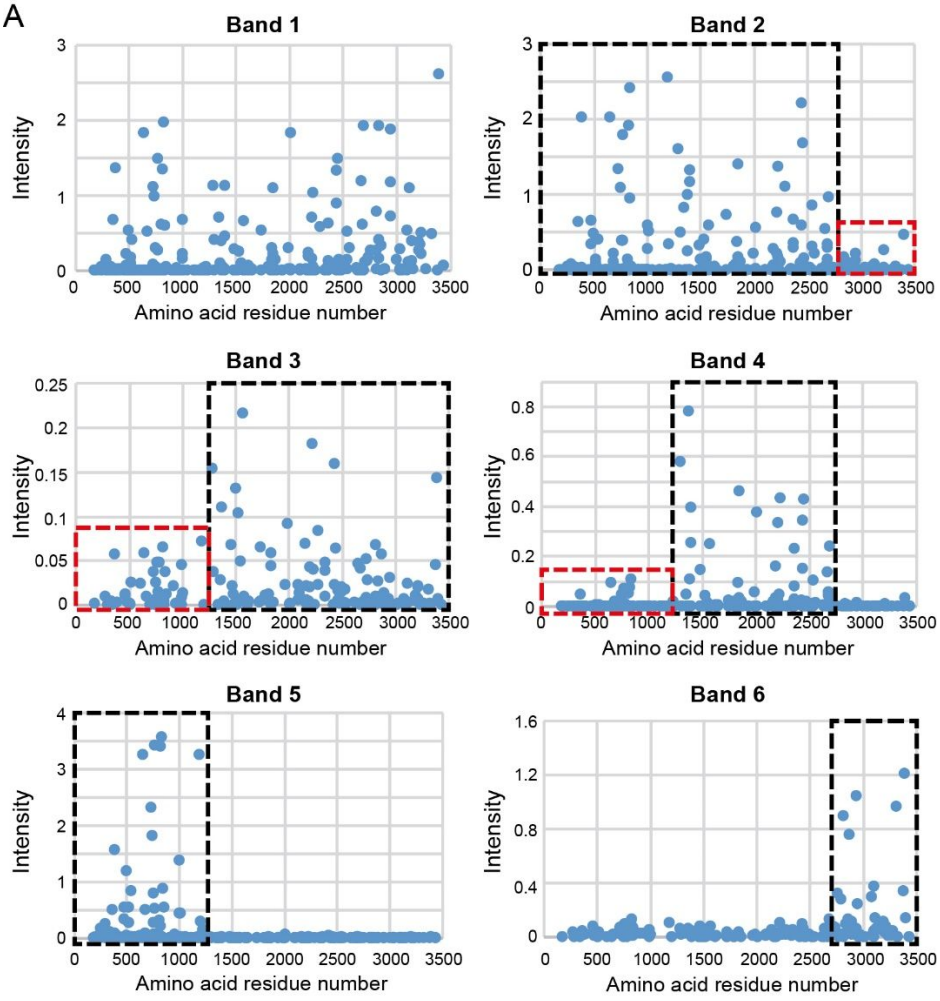


Supplementary Figure 6. Image of complete SDS-PAGE gel that is shown in part in Figure 2C.



Supplementary Figure 7. LC-MS/MS analysis of purified Reelin proteins. (A) Bands 1 to 6 in Figure 2C were individually analyzed by LC-MS/MS. The relative intensities of the identified Reelin peptides (blue dots) with start positions at the indicated amino acid residue numbers are presented for each band. The black boxes indicate the regions in which peptides with relatively high intensities are frequently identified. The red boxes indicate the peptide positions with relatively low but significant intensity that likely resulted from minor contaminations in SDS-PAGE and/or multiple runs of LC-MS/MS analysis. (B) The names of the proteins that were identified, and the number of times they were detected, are shown for Band 3.

A



B

Protein	Number of peptides	Total spectral counts
sp|P78509|RELN_HUMAN	133	567
sp|P04264|K2C1_HUMAN	45	153
sp|P13645|K1C10_HUMAN	29	80
sp|P35527|K1C9_HUMAN	34	75
sp|P35908|K22E_HUMAN	26	68
sp|Q86YZ3|HORN_HUMAN	11	21

Supplementary Figure 8. Reelin-induced Dab1 phosphorylation. Representative images used for optical density measurements of phosphorylated Dab1(Tyr^{232})-positive SH-SY5Y cells following treatment with purified Reelin-f18 or commercial rReelin-CF. Scale bar: 50 μm.