Supplemental Data

A recurrent *COL6A1* pseudoexon insertion causes muscular dystrophy and is effectively targeted by splice-correction therapies

Supplemental Acknowledgments

***COL6A1* Intron 11 Study Group:** Enrico S. Bertini[1], Mary-Lynn Chu[2], James Collins[3], Giacomo Comi[4], Sidney M. Gospe, Jr[5,6], Carla Grosmann[7], Janbernd Kirschner[8], Brian D. Kossak[9], Baiba Lace[10], Meganne E. Leach[11], Edward Leung[12], Luciano Merlini[13], Pinki Munot[14], Andrés Nascimento Osorio[15], Yoram Nevo[16], Ishizo Nishino[17], Susana Quijano-Roy[18], Randal Richardson[19], Monique M. Ryan[20,21], Mena Scavina[22], Mordechai Shohat[16], Tanya Stojkovic[23], Mónica Troncoso[24], and Grace Yoon[25]

[1]Unit of Neuromuscular and Neurodegenerative Disorders, Laboratory of Molecular Medicine, Department of Neurosciences, Bambino Gesù Children's Hospital, IRCCS, Rome, Italy.

[2]Department of Neurology, NYU-School of Medicine, New York, USA.

[3]Neurology Division, Cincinnati Children's Hospital Medical Center, Cincinnati, USA.

[4]Dino Ferrari Centre, Neuroscience Section, Department of Pathophysiology and Transplantation, Neurology Unit, IRCCS Foundation Ca'Granda Ospedale Maggiore Policlinico, University of Milan, Milan, Italy.

[5]Departments of Neurology and Pediatrics, University of Washington, Seattle, USA.

[6]Seattle Children's Hospital, Seattle, USA.

[7]Division of Neurology, Rady Children's Hospital, San Diego, USA.

[8]Division of Neuropediatrics and Muscle Disorders, University Medical Center, Freiburg, Germany.

[9]Pediatric Neurology, Children's Hospital at Dartmouth, Manchester, USA.

[10]Centre Hospitalier Universitaire de Québec, Québec City, Canada.

[11]Neuromuscular and Neurogenetic Disorders of Childhood Section, Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD 20892 USA.

[12]Section of Pediatric Neurology, The Children's Hospital Research Institute of Manitoba, Winnipeg, Canada.

[13]School of Physical and Rehabilitative Medicine, University of Bologna, Bologna, Italy.

[14]Dubowitz Neuromuscular Centre, UCL Great Ormond Street Institute of Child Health, London, United Kingdom.

[15]Neuromuscular Unit, Neuropediatrics Department, Hospital Sant Joan de Déu, Barcelona, Spain.

[16]Institute of Neurology, Schneider Children's Medical Center of Israel, Petach Tikvah, Israel

[17]Department of Neuromuscular Research, National Center of Neurology and Psychiatry, Tokyo, Japan.

[18]Garches Neuromuscular Centre (GNMH), Raymond Poincaré University Hospital (UVSQ), Garches, France.

[19]Gillette Children's Specialty Healthcare, St. Paul, USA.

[20]Department of Neurology, Royal Children's Hospital Melbourne, Parkville, Australia.

[21]Murdoch Children's Research Institute, Parkville, Australia.

[22]Nemours/Alfred I. DuPont Hospital for Children, Wilmington, USA.

[23]Institute of Myology and AP-HP, Groupe Hospitalier La Pitié-Salpêtrière, Centre de référence des maladies neuromusculaires, Paris, France.

[24]Pediatric Neurology, Hospital San Borja Arriarán, Santiago, Chile.

[25]Clinical and Metabolic Genetics, The Hospital for Sick Children, Toronto, Canada.

Supplemental Methods

Cloning (shotgun and splicing reporters)

Shotgun cloning of reverse-transcriptase PCR (RT-PCR) products generated using cDNA from cultured fibroblasts or muscle tissue was carried in pCR4-Blunt-TOPO following manufacturer's instructions (Invitrogen/ThermoFisher Sci), and transformed in chemically-competent TOP10 bacteria (Invitrogen/ThermoFisher Sci). To screen, a series of positive colonies were randomly selected, from which DNA was prepared (Zymo Research), and sent for sequencing (Genewiz).

Minigene (splicing reporters) were prepared by cloning sequences from human *COL6A1* gene into the Exontrap pET01 vector (MoBiTec GmbH, Göttingen, Germany). The template genomic DNA was from patient US2, and was amplified using the primer pairs IVS-11, Ex-11-13, or Ex-10-13 (Supplemental Table 4), that included XhoI (forward) and BamHI (reverse) sites. All primers included either an XhoI (forward) or BamHI (reverse) site. PCR was carried out in a 20 µL total volume using 70 ng of gDNA, 0.5 µM of each primer, and a final concentration of 1X of the Phusion High-Fidelity PCR master mix with HF Buffer (New England Biolabs (NEB), Ipswich, MA). Amplification was performed using a touchdown protocol. Following the initial denaturation at 98˚C for 30 sec, the first 2 cycles were run at 98˚C for 10 sec, 70˚C for 20 sec, and 72˚C for 30 sec. The following cycles were run under the same conditions, with the exception of the annealing temperature that was decreased to 68˚C for 2 cycles, 66˚C for 4 cycles, 64˚C for 5 cycles, 62˚C for 5 cycles, 60˚C for 5 cycles, 58˚C for 5 cycles, and 56˚C for 5 cycles. The final elongation was 72˚C for 10 min. The purified PCR products (QIAquick PCR Purification kit, Qiagen, Hilden, Germany), as well as the vector (1-1.5 µg) were double digested

with 40 U of XhoI (NEB) and 40 U of BamHI-HF (NEB) in Cutsmart buffer (NEB) for 45 minutes at 37˚C, then gel purified (QIAquick Gel Extraction kit, Qiagen). Ligation of vector (~50ng) with each PCR product (~50 ng) was carried out at room temperature for 45 min using 400U of T4 DNA Ligase (NEB), in a 1X final concentration of T4 DNA Ligase buffer (NEB). Two ul was used for transformation in 50 µL of chemically-competent bacteria (Max Efficiency DH5 α , Invitrogen/ThermoFisher Scientific (ThermoFisher Sci), Waltham, MA), then spread on ampicillin agar plates. After positive colonies were mini-preped (Zyppy Plasmid Miniprep kit, Zymo Research, Irvine, CA) and verified by sanger sequencing (Genewiz), larger yields were prepared using the HiSpeed Plasmid Maxi kit (Qiagen). For each plasmid, two clones were selected: one that carried the wildtype c, and one that carried the mutant (T) genotype at the c.930+189 locus. Sanger sequencing further validated that no variant other than the C>T mutation was present in the remaining of the sequence.

DNA/cDNA Amplification and Sequencing

For minigene-expressing cell samples, cDNA was amplified using the following primers: 5'-ggattcttctacacaccc-3'and 5'-ccgggccacctccagtgcc-3', using either the Taq Advantage 2 (Clontech/Takara Bio USA, Inc., Mountain View, CA) or the KAPA (Kapa Biosystems, Wilmington, MA) polymerases. For the Taq Advantage 2 polymerase reaction, 2 µL of cDNA was mixed in a final volume of 20 µL with 0.2 mM of dNTPs, 0.3 µM of each primer, and final concentrations of 1X of buffer and 0.75X of Taq Advantage 2 enzyme, and was run at 95˚C for 1 min, followed by 5 cycles of 95˚C for 30 sec, and 68˚C for 30 sec, 5 cycles of 95˚C for 30 sec, 64˚C for 30 sec, and 68˚C for 30 sec, 18 cycles of 95˚C for 30 sec, 55˚C for 30 sec, and 68˚C for 30 sec, and a final extension of 68˚C for 1 min. For the KAPA polymerase reaction, 0.5 µL of

cDNA was mixed with in a final volume of 20 µL with 0.3 mM of dNTPs, 0.3 µM of each primer, 0.4 U of KAPA HiFi HotStart and a final concentration of 1 X of buffer. Cycling conditions were: 95˚C for 5 min, followed by 35 cycles of 98˚C for 20 sec, 64˚C for 15 sec, and 72˚C for 1 min. Products were run on a 2.5% agarose gel and imaged using the ChemiDoc XRS+ Imaging System (Bio-Rad, Hercules, CA). Lanes were plotted using ImageJ to measure the area under the curve (AUC). Within each lane, the AUC of the pseudoexon-included amplicon was divided by the AUC of the wildtype amplicon to obtain the Pseudoexon/WT ratio.

To detect pseudoexon expression in skin fibroblasts and muscle biopsy samples, 2 ul of cDNA was amplified, using the 2X KAPA HiFi HotStart ReadyMix 2X master mix system (Kapa Biosystems) at a final concentration of 1X, with 0.3 µM of each primer (Supplemental Table 4) in a total volume of 20 µL. Cycling protocol was as described above (for the KAPA HiFi HotStart), using a 35 cycle-protocol for fibroblast-derived samples, and a 40-cycle protocol for muscle-derived samples.

To sequence specific cDNA isoforms, RT-PCR products were run on low percentage agarose gels, and fragments were isolated using Gel Extraction Kit (Qiagen). PCR products were sent to Genewiz for sequencing.

For PCR of genomic DNA (gDNA) post Cas9 or Cas9+U1D1 transfection and nucleofection of human patient dermal fibroblasts, the primers used are listed in Supplemental Table 4.

Percent pseudoexon expression

The percent pseudoexon expression was determined from semi-quantitative RT-PCR gel images using ImageJ. After acquisition, images were exported as tagged image file format (.tiff), and transferred to ImageJ. Background was subtracted, then each lane was identified with the rectangular selection tool, and the plot lane tool was used to visualize the peak profiles of each lane. The area under the curve (AUC) for each peak was measured, and used to calculate the percent pseudoexon expression ($AUC_{pseudoexon}/(AUC_{wildtype}+AUC_{pseudoexon})$).

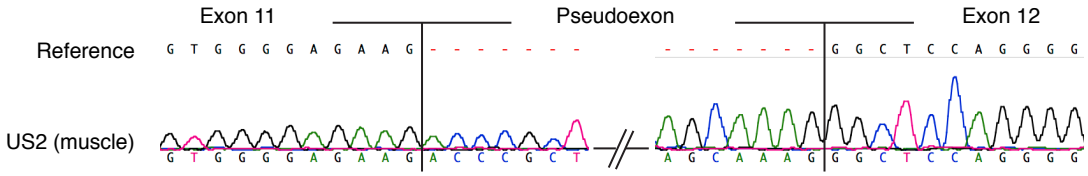
Liquid Chromatography-tandem Mass Spectrometry (LC-MS/MS)

The SDS-PAGE band containing collagen VI complex was excised, treated with TCEP (Sigma-Aldrich) and alkylated with NEM (Sigma-Aldrich). Half of the gel band was digested with trypsin (Promega) overnight at 37 °C. The other half of the band was digested with chymotrypsin (Roche Diagnostic) overnight at 25 °C. LC/MS/MS analysis was performed using an Orbitrap Elite mass spectrometer (Thermo Fisher Sci) coupled with a 3000 Ultimate high-pressure liquid chromatography instrument (Thermo Fisher Sci). Peptides were separated over a 45-min gradient on a ES802 column (Thermo Fisher Sci). The LC-MS/MS data were acquired in data-dependent mode. The resolution of the survey scan (300-1600 m/z) was set at 60k at m/z 400 with a target value of 10×10^6 ions. Collision-induced dissociation (CID) were performed on the top ten most abundant precursor ions with an isolation window of 2.0 Da.

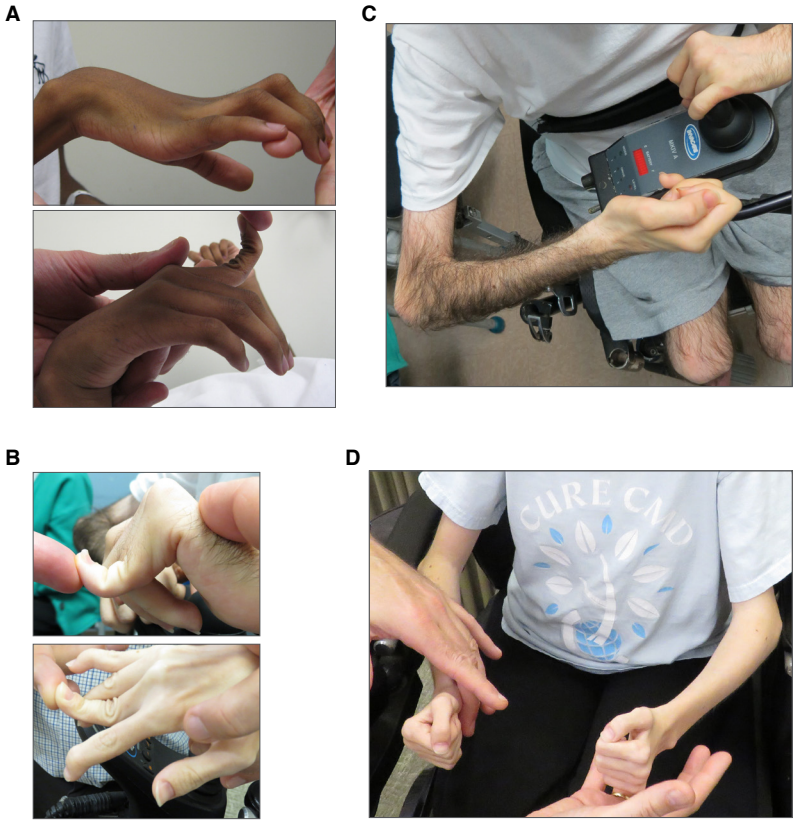
Database search was performed using Mascot Daemon (2.5.0) (Matrix Science) against a house-built database which contains the collagen VI sequence with peptide insertion and sequences from NCBI Human database. The error tolerances for precursor and product ions were ±5 ppm and ±0.3 Da, respectively. Nethylmaleimide (Cys) and Oxidation (Met) were set as variable

modifications. Semitrypsin and None were chosen as enzyme for data obtained from trypsin and chymotrypsin digestion, respectively. All spectra of peptides matched to the insertion region were manually checked.

Supplemental Figures

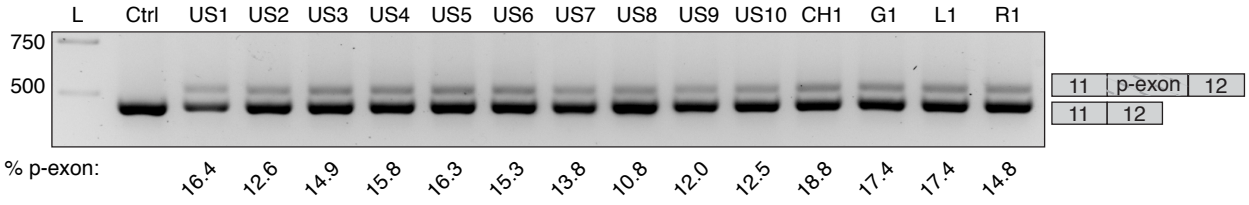


Supplemental Figure 1. Sequencing validates the presence of a 72-bp pseudoexon insertion between *COL6A1* exons 11 and 12. Sanger sequencing chromatogram of gel-separated cDNA amplicons from muscle tissue shows the boundaries of the 72-bp pseudoexon sequence inclusion.

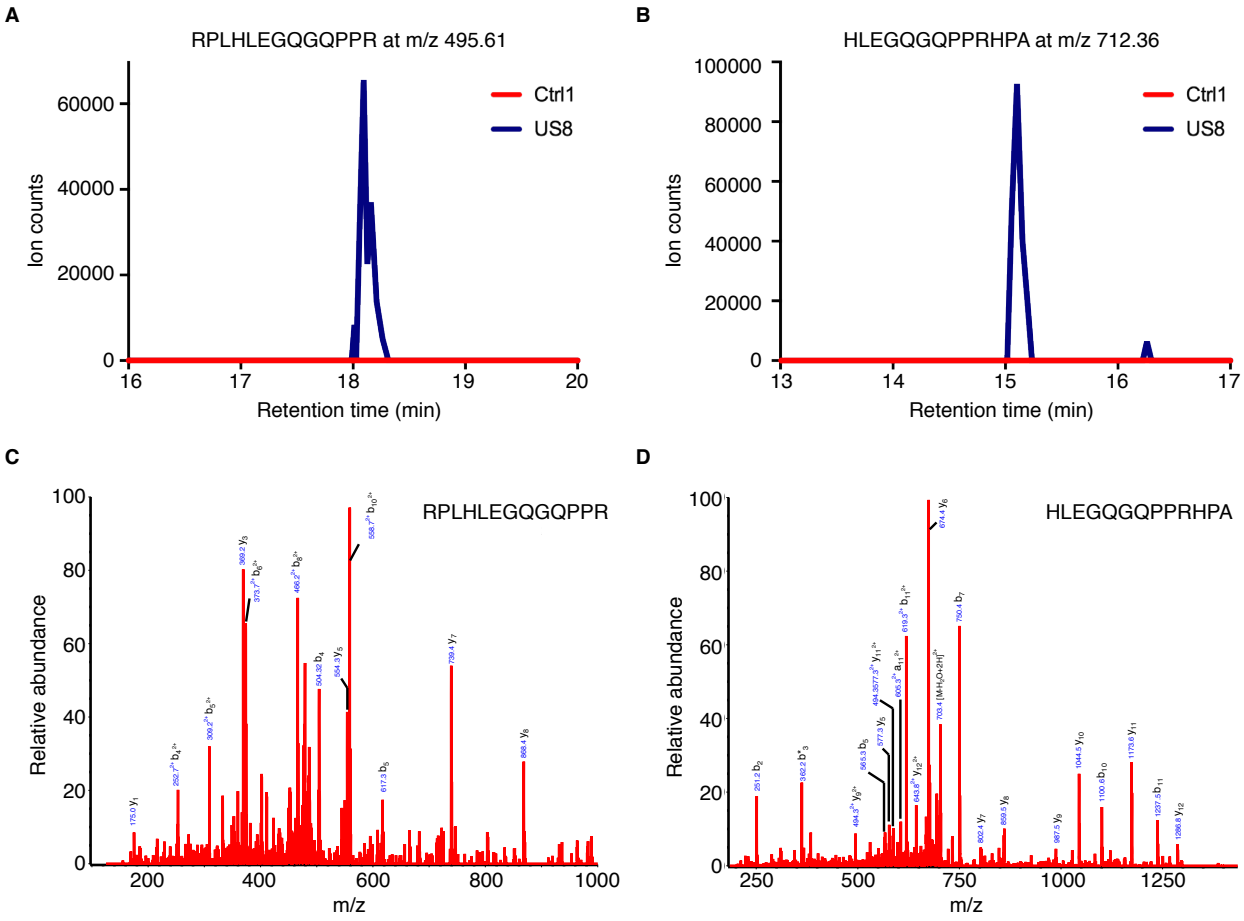


Supplemental Figure 2. The *COL6A1* +189C>T mutation manifests as severe COL6-RD.

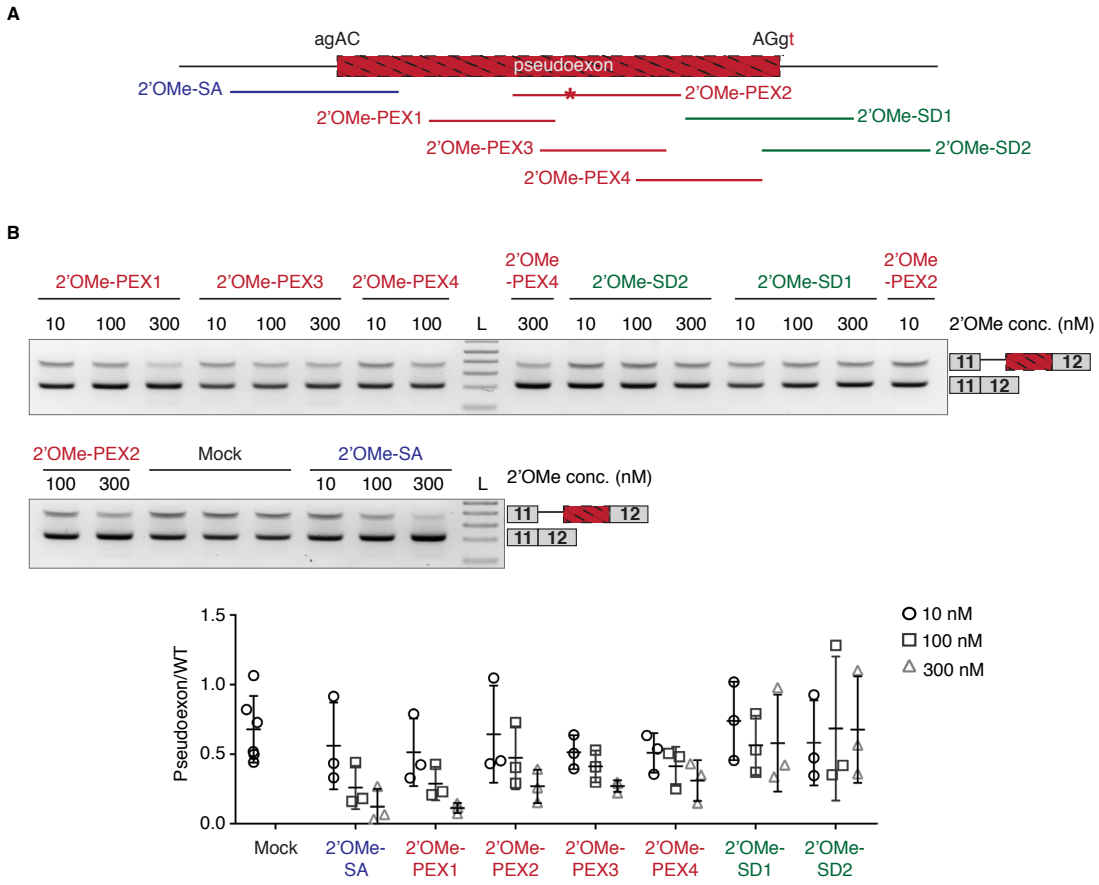
Photographs of *COL6A1* +189C>T patients showing (**A** and **B**) distal joint hypermobility, and (**C** and **D**) proximal joints contractures.



Supplemental Figure 3. Expression of the *COL6A1* pseudoexon in cultured patient-derived dermal fibroblasts. cDNA amplicons from dermal fibroblasts, generated using primers spanning *COL6A1* exons 10 to 20, showing low expression levels of the pseudoexon in all cultured fibroblasts assayed. PCR fragments densities were determined by gel quantification (ImageJ), and used to calculate the percent pseudoexon expression (% p-exon), reported as the average of three quantifications (below). Gel image is representative of three amplifications.



Supplemental Figure 4. Detection of the pseudoexon peptide in patient-derived dermal fibroblasts by Liquid Chromatography Mass Spectrometry. (**A** and **B**) Extracted ion chromatogram (XIC) of peaks related to fragments from the pseudoexon peptide in the LC-MS/MS runs of samples digested with trypsin (**A**) or chymotrypsin (**B**). Unique fragments from the pseudoexon peptide was detected only in US8 samples, not in Ctrl1 samples. (**C** and **D**) MS/MS spectra for RPLHLEGQGQPPR (**C**) and HLEGQGQPPRHPA (**D**).



Supplemental Figure 5. Screening of 2'O-Methyl antisense oligomers for pseudoexon skipping in minigene-expressing cells. (**A**) Location of the 2'-O-Methyl (2'OMe) antisense oligomers targeting either the splice acceptor (SA), the splice donor (SD), or sequences within the pseudoexon (PEX). Asterisk in 2'OMe-PEX2 denotes the presence of a mismatch between the oligomer and the target sequence. (**B**) Screening of the 2'OMe oligomers in pET01 Ex-11-13-expressing cells. Agarose gels of RT-PCR products amplified as in Fig. 1E (top), representative of triplicate transfections. PCR fragments densities were quantified (ImageJ), and reported as the ratio of pseudoexon over normal (Pseudoexon/WT) amplicons (bottom). Bars represent the average ratio±standard deviation of triplicate transfections. Mock=transfection reagent only.

Supplemental Table 1. Phenotypic characteristics of the *COL6A1* +189C>T cohort.

PATIENT IDENTIFIER	SEX	RACE/ETHNICITY	AGE AT MOST RECENT CLINICAL ASSESSMENT (YEARS)	AGE STOPPED WALKING/STARTED USING WHEELCHAIR FULL TIME (YEARS)	AGE STARTED BIPAP (YEARS)
US1	Female	Hispanic/Mexican	8	6	8
US2	Male	Caucasian/Ashkenazi Jewish	29	11	18
US3	Male	Hispanic	15	3.5	14
US4	Female	Caucasian	30	12	21
US5	Female	Caucasian	24	10	21
US6	Female	Caucasian	23	13	14
US7	Female	Caucasian	15	5	13
US8	Female	Hispanic	17	10	
US9	Male	African American	15	10	15
US10	Male	Hispanic/Guatemalan and Costa Rican	8	still ambulant	8
US11	Male	Hispanic	10	8	11
CA1	Female	Asian/Indian	12	NA	12
CA2	Male	Asian/Bengali	13	9	13
CA3	Male	Caucasian	28	9	12
CH1	Female	Hispanic/Chilean	24	10	21
G1	Male	Caucasian/German	19	10	not yet
IR1	Male	Caucasian/Irish	10	9	11
L1	Male	Caucasian/Latvian	11	8	not yet
R1	Female	Caucasian/Romanian	5	NA	NA
UK1	Male	Caucasian/British	7	still ambulant	not yet
UK2	Male	Caucasian/Irish	7	still ambulant	not yet
UK3	Female	Caucasian/British	38	13	15
UK4	Female	Caucasian/Greek	16	8	8
UK5	Female	Caucasian/Spanish	5	still ambulant	not yet
F1	Male	Caucasian	11	6	not yet
F2	Female	Caucasian/French and African/Algerian	22	14	refused
IT1	Female	Caucasian/Italian	7	6	NA
IT2	Female	Caucasian/Italian	11	4	11
IT3	Male	Caucasian/Italian	9	5	9
IT4	Female	Caucasian/Italian	7	still ambulant	7
IT5	Male	Caucasian/Italian	17	12	14
A1	Female	Caucasian/Australian	12	8	11
S1	Male	Caucasian/Spanish	7	still ambulant	not yet
S2	Female	Caucasian/Spanish	8	still ambulant	not yet
IS1	Female	Caucasian/Middle Eastern	5	still ambulant	not yet

Supplemental Table 2. Sequencing results of the shotgun cloning of reverse-transcriptase PCR products derived from US14 dermal fibroblasts.

rs1980982 genotype	Pseudoexon insertion (+/-)	Number of clones
C	-	11
C	+	0
T	-	10
T	+	6

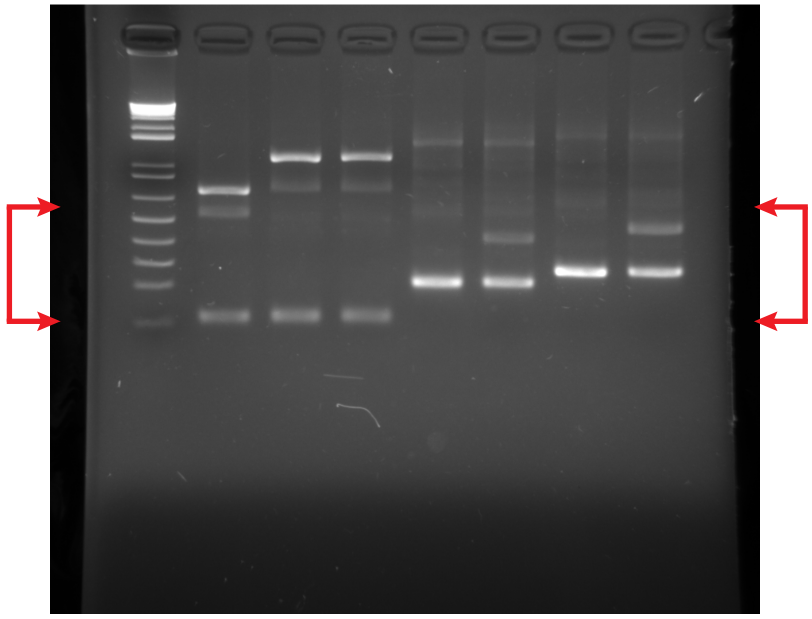
Supplemental Table 3. Sequences of antisense oligomers.

Oligomer ID	Oligomer sequence
PMO-SA	AGGACACCTGGTCTCCTGTCAGGGA
PMO-PEX1	GCTGTCCTTGTCCTTCCAGATGGAG
PMO-PEX2	GTGGCTGTCCTTGTCCTTCCAGATG
PMO-PEX3	GTGCCTGGGTcGCTGTCCTTGTCCT
PMO-PEX4	TTGTCCTTCCAGATGGAcGGGAC
PMO-SD	TGAAAGTGACACAGGCACCTTTGCT
2'OMe-SA	GUGGAGCGGGUCUGCAGGACACGUG
2'OMe-PEX1	AGAUGGAGGGGACGGCGAGG
2'OMe-PEX2	GGCUGUCCUUGUCCUcCCAGAUGGA
2'OMe-PEX3	GGCUGUCCUUGUCCUUCCAG
2'OMe-PEX4	GUGCCUGGGUGGCUGUCCUU
2'OMe-SD1	AGGCACCUUUGCUGGGUGCCUGGGU
2'OMe-SD2	UGAAAGUGACACAGGCACCUUUGCU

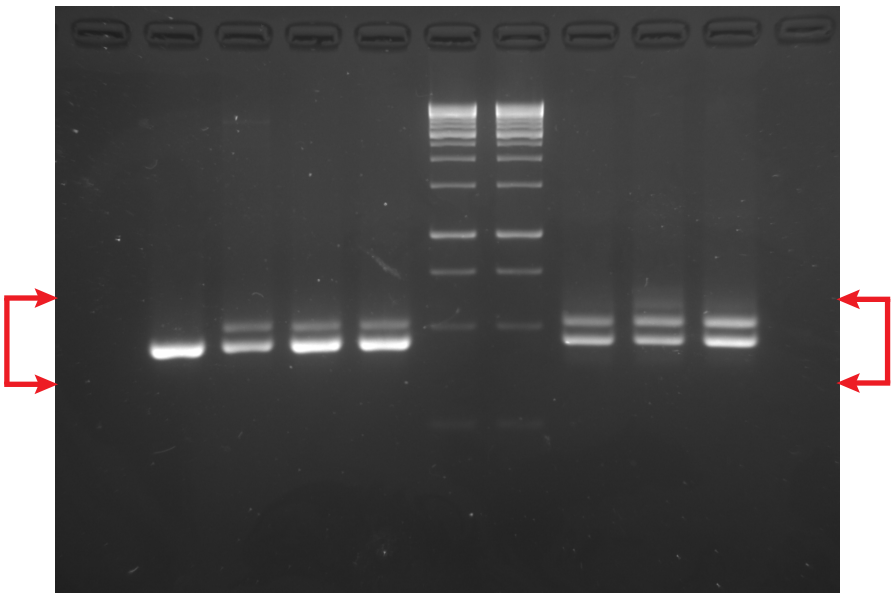
Supplemental Table 4. Sequences of primers used for cloning, amplification, and sequencing.

Genomic DNA amplification			
	Primer forward	**Primer reverse**	
COL6A1 +189C>T diagnostic testing	tgttgggtaccagggaatgaaggt	aaacgaaggcaggagtcaga	
Cas9 cutting efficiency	gggtaccagggaatgaaggt	gaggggacaacatcactgct	
Reverse-transcriptase PCR (RT-PCR)			
	Primer forward	**Primer reverse**	
Minigene constructs	ggattcttctacacaccc	ccgggccacctccagtgcc	
COL6A1 exons 10 to 15	acctgttgggtaccagggaatgaa	ttctcctttcagtccaaaggcacc	
COL6A1 exons 10 to 20	acctgttgggtaccagggaatgaa	accagggtctcctcttggtc	
Quantitative reverse-transcriptase PCR (qRT-PCR)			
	Primer forward	**Primer reverse**	**Probe**
COL6A1 pseudoexon-specific transcripts	taccagggaatgaagggaga	cctggagccctttgctg	atctggaaggacaaggacagccac
Total *COL6A1* transcripts	ccgactgcgctatcaagaa	aatcaggtacttattctccttcaggt	Roche UPL #17
PGK1	cagctgctgggtctgtcat	gctggctcggctttaacc	Roche UPL #67
Plasmid cloning and sequencing			
	Primer forward	**Primer reverse**	
Minigene IVS-11	ctactcgaggagtgaggctcgacctcgga	gatggatccgcggagacaggaagcgggtg	
Minigene Ex-11-13	ctactcgaggtggtgccctcagccttg	gatggatccctccccgtccatcctatctc	
Minigene Ex-10-13	ctactcgagcctctctcggcctgacca	gatggatccctccccgtccatcctatctc	
guideRNA Upstream (U1)	caccggacacgtggtctcctgtca	aaactgacaggagaccacgtgtcc	
guideRNA Downstream (D1)	accgggtggggtgaaagtgacacgt	taaaacgtgtcactttcaccccacc	
U6	gactatcatatgcttaccgt		
BGH PA	tgcatcgcattgtctgagtagg		

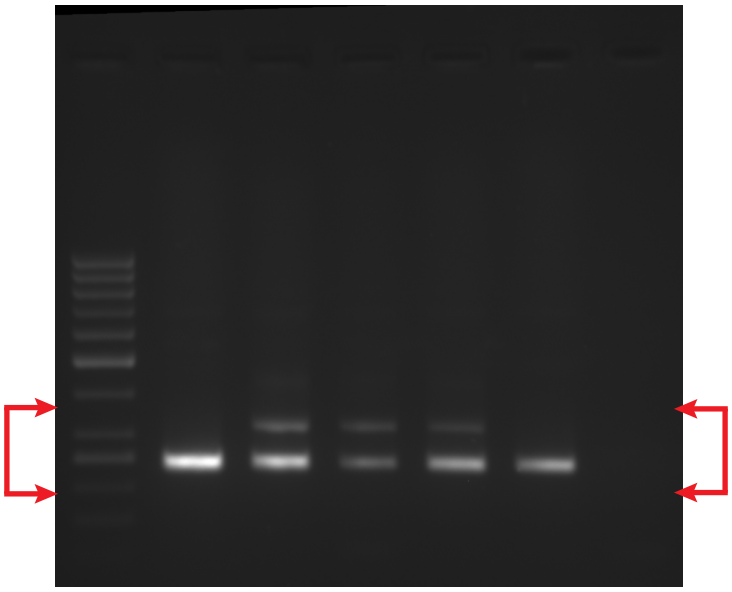
1E



1G

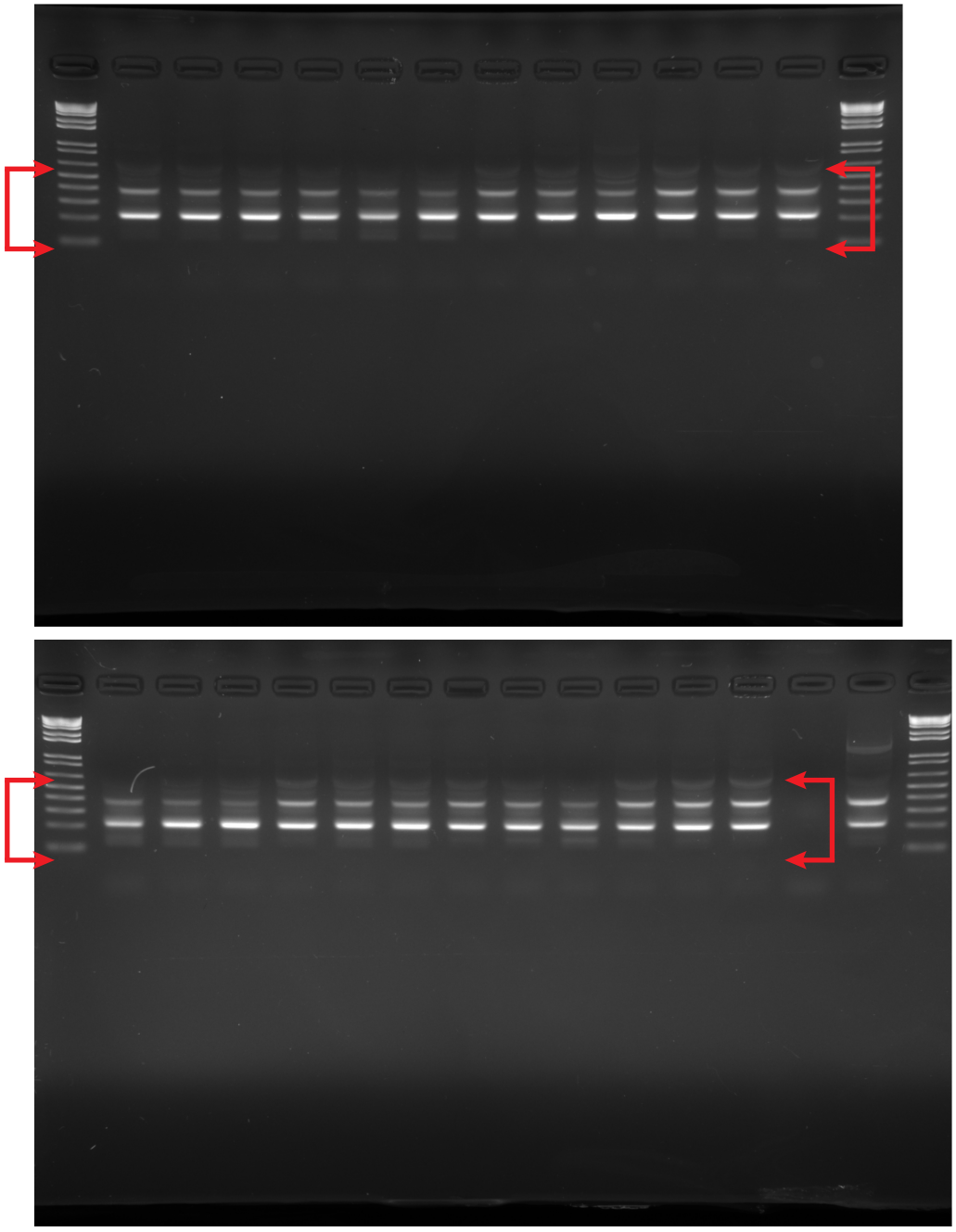


1H

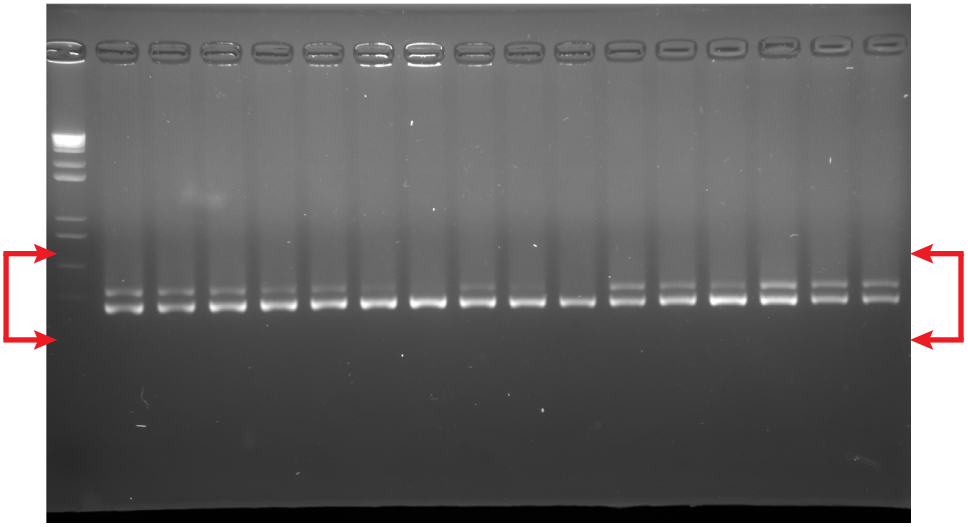


Full unedited gels for Figure 1

3B



3C

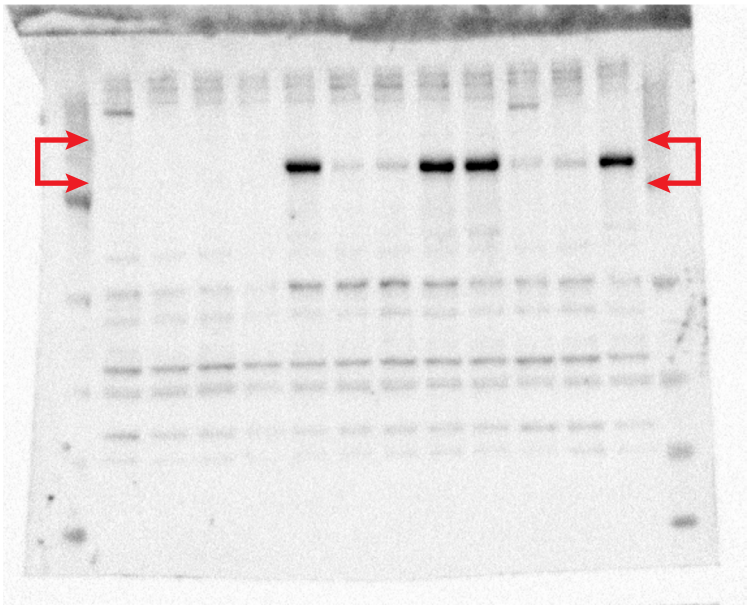


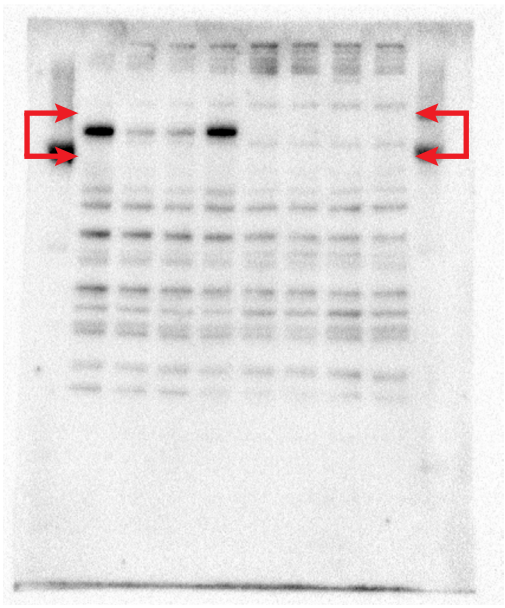
Full unedited gels for Figure 3

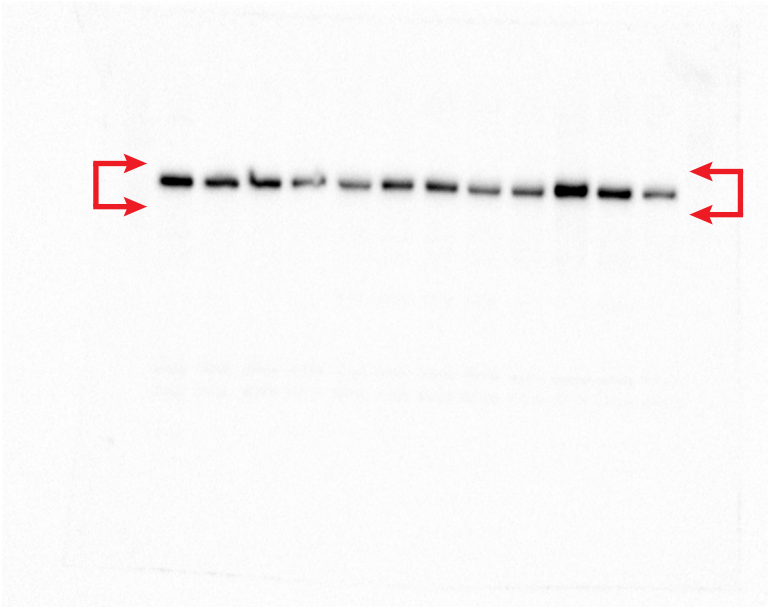
4A

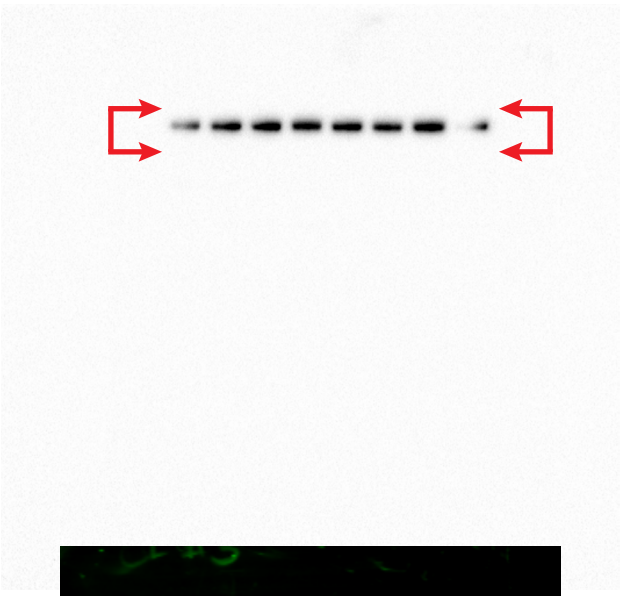
Top left panels

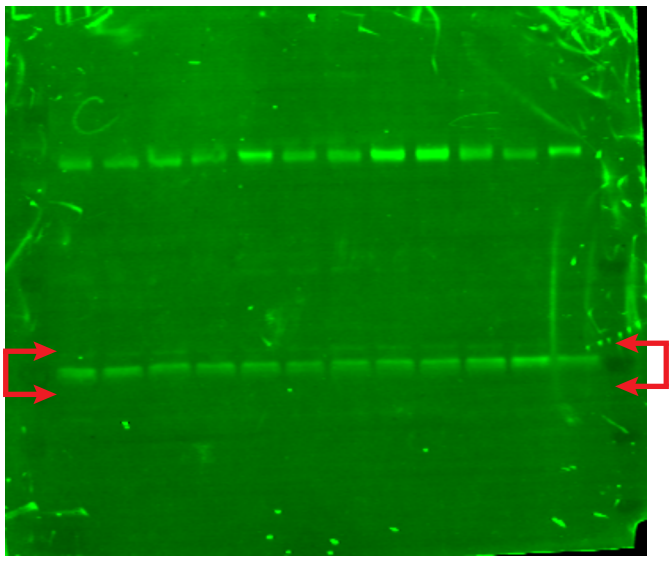
Top right panels

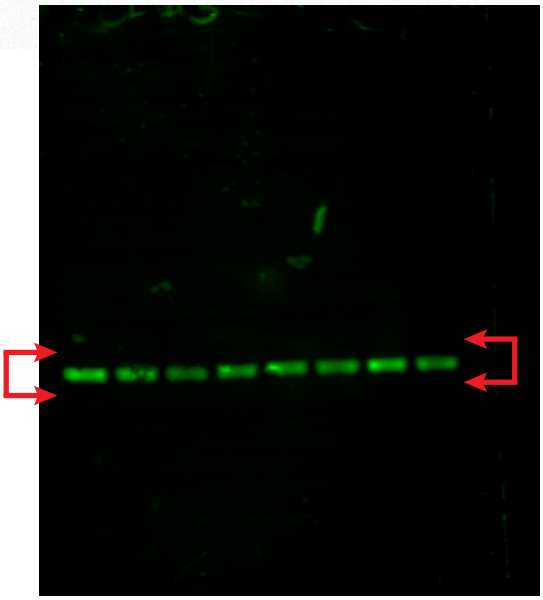










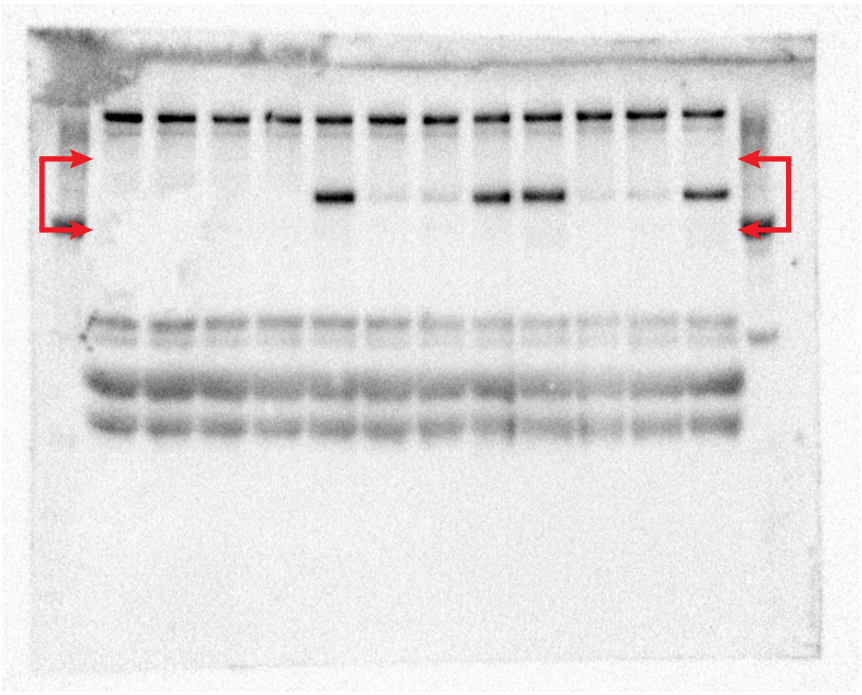


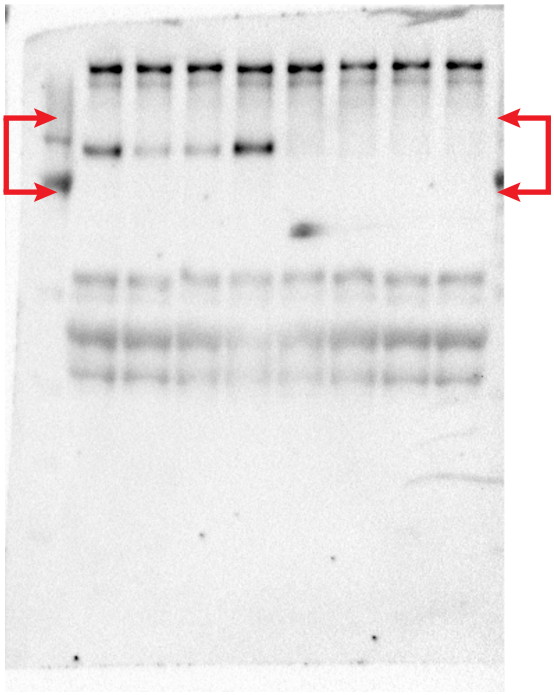
Full unedited gels for Figure 4 (page 1 of 2)

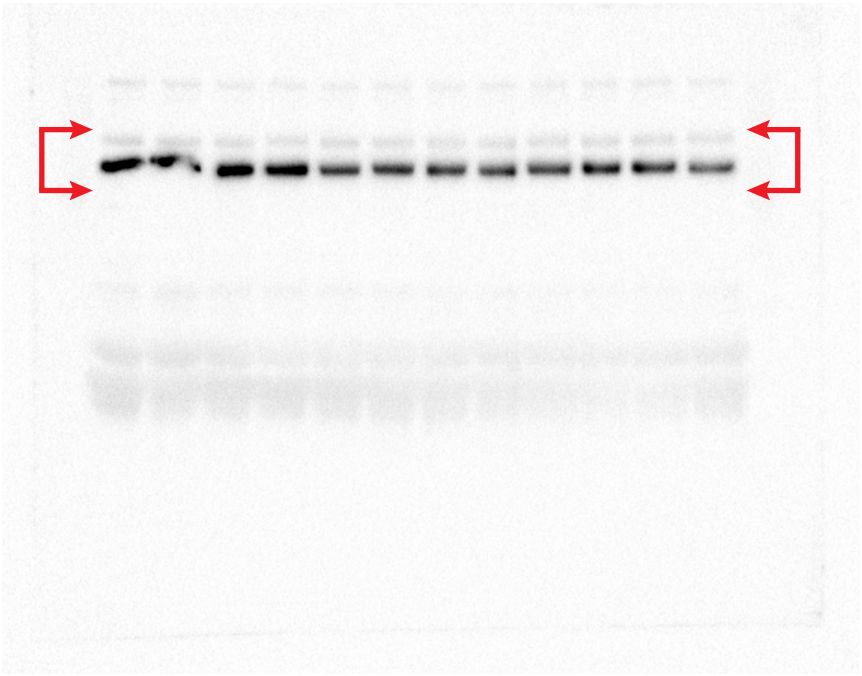
4A

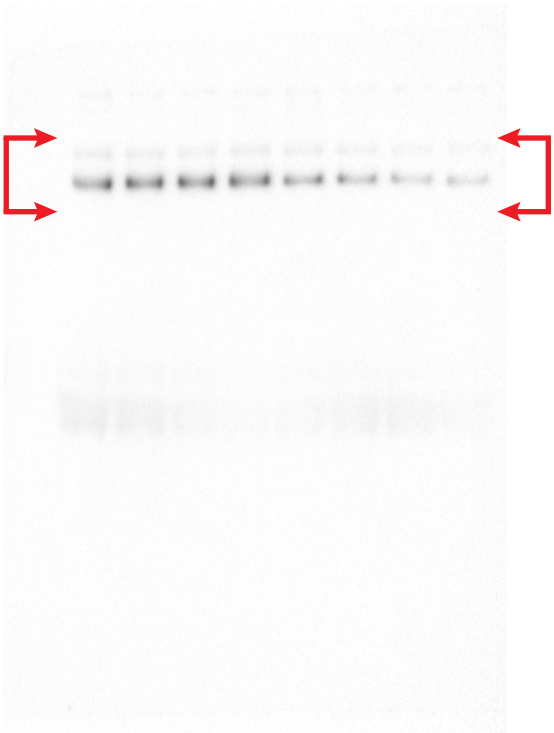
Bottom left panels

Bottom right panels



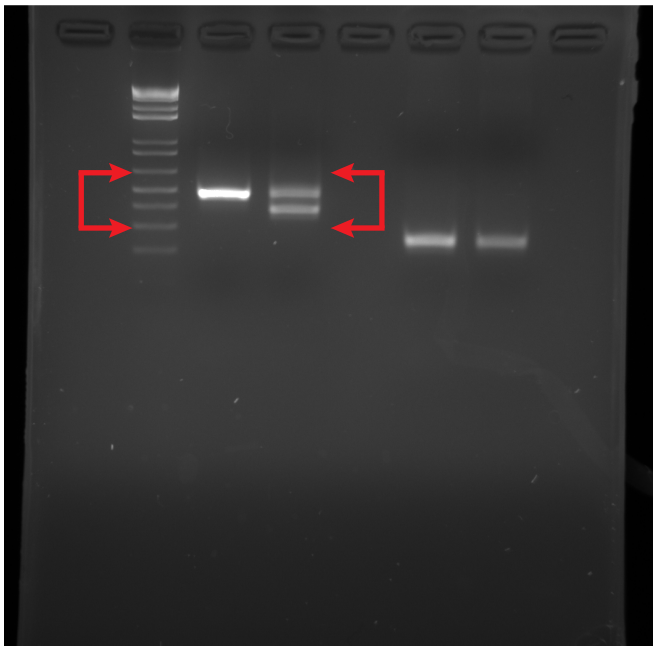




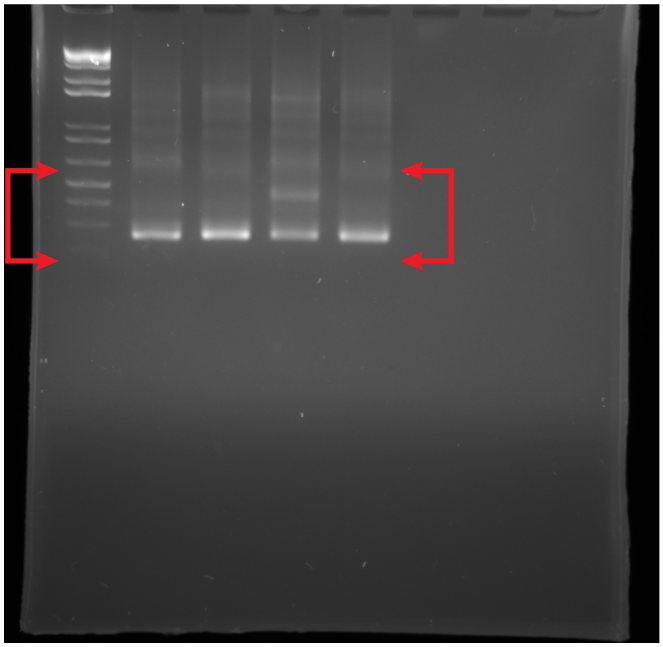


Full unedited gels for Figure 4 (page 2 of 2)

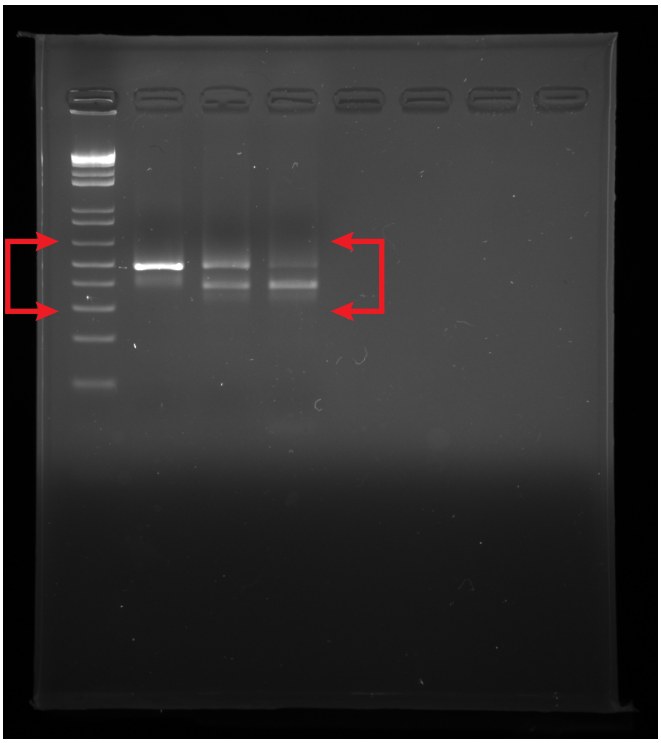
5B



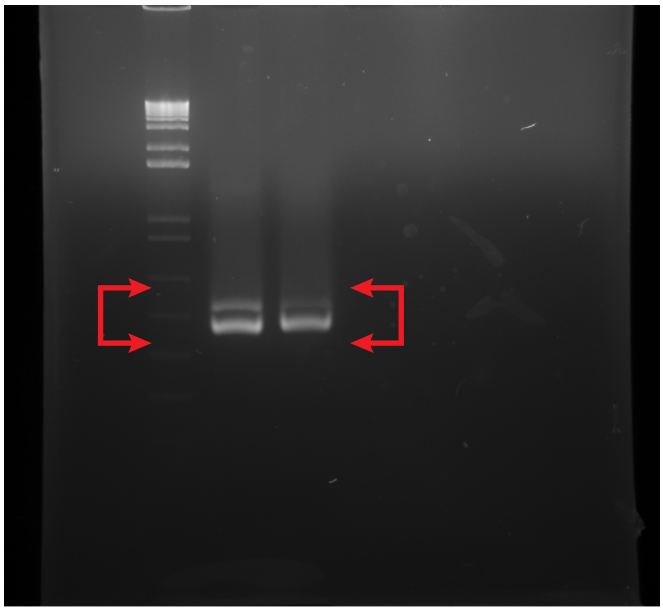
5C



5D



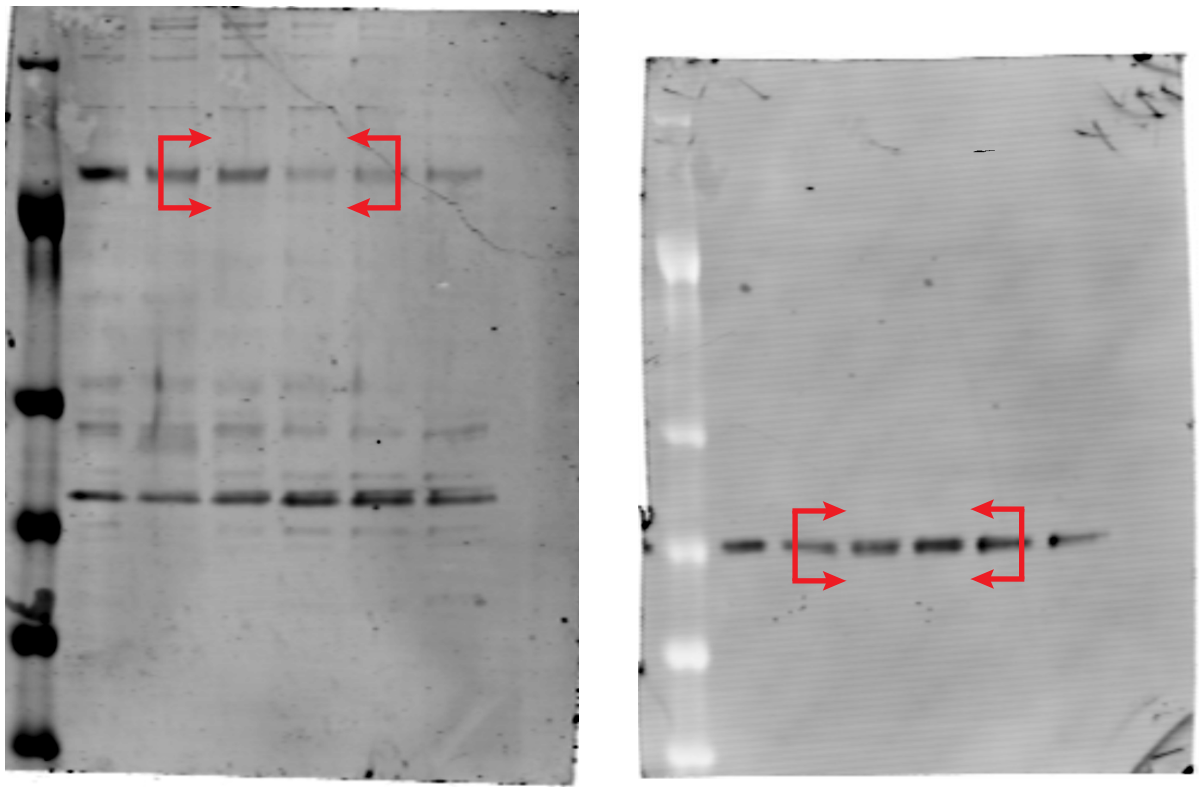
5E



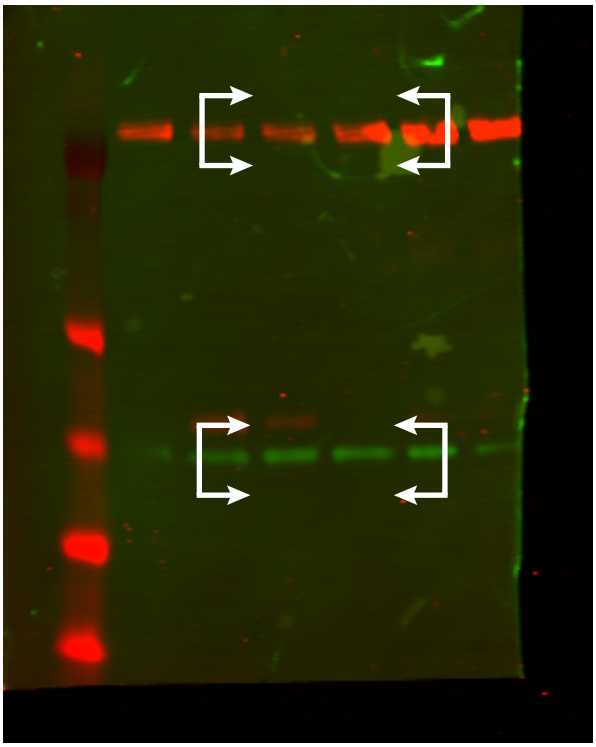
Full unedited gels for Figure 5 (B-E)

5F

Left Top panels



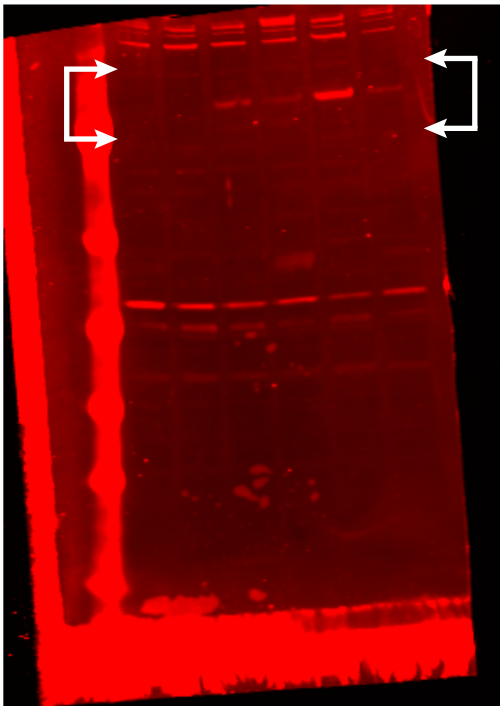
Left Bottom panels

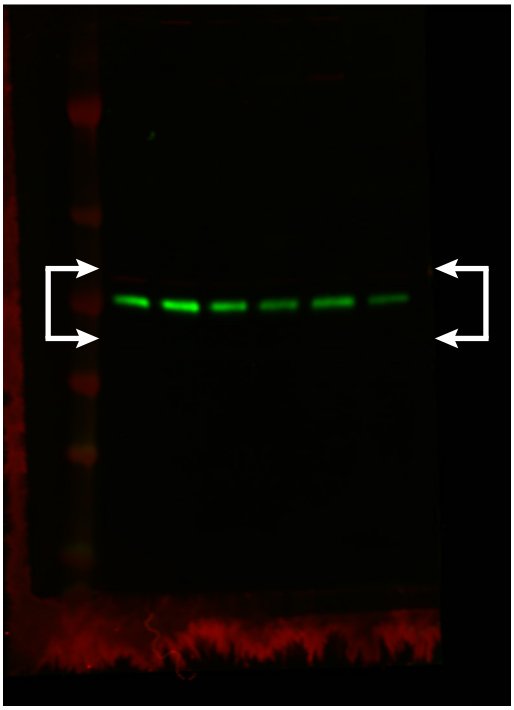


Full unedited gels for Figure 5 (F)

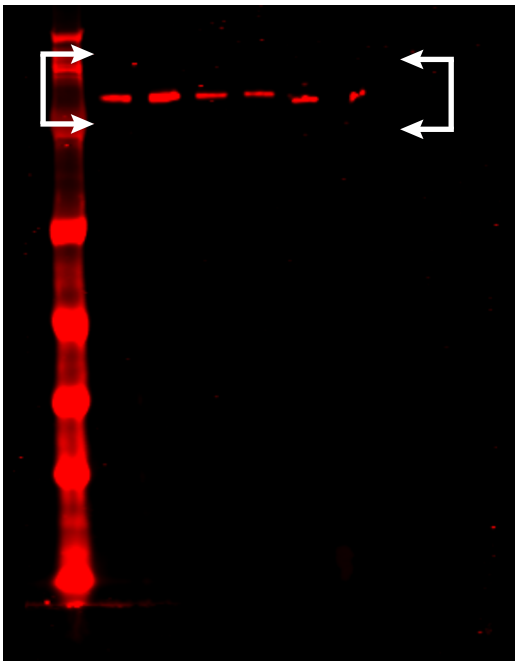
5F

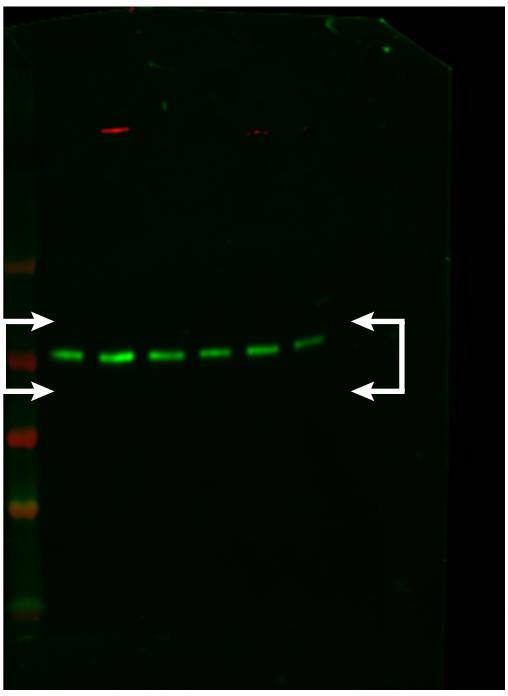
Right Top panels

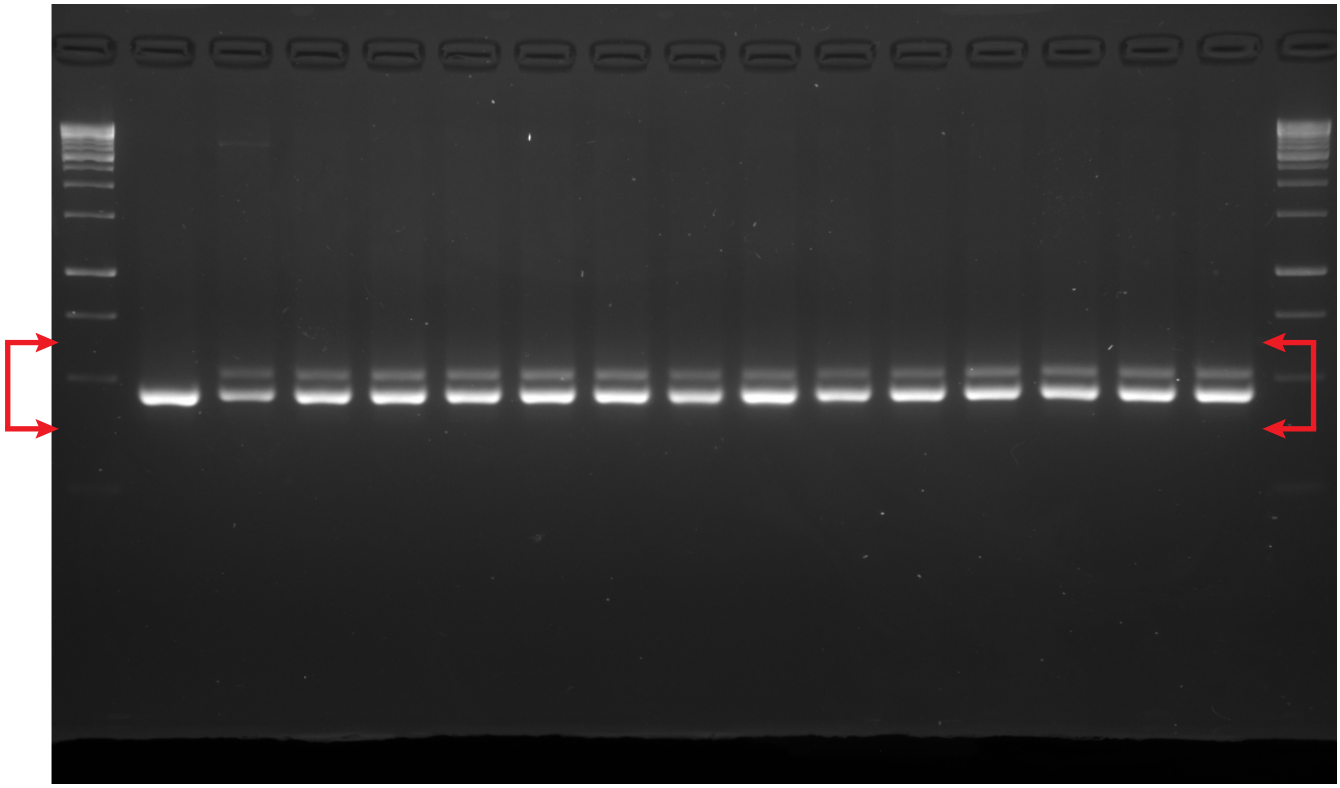




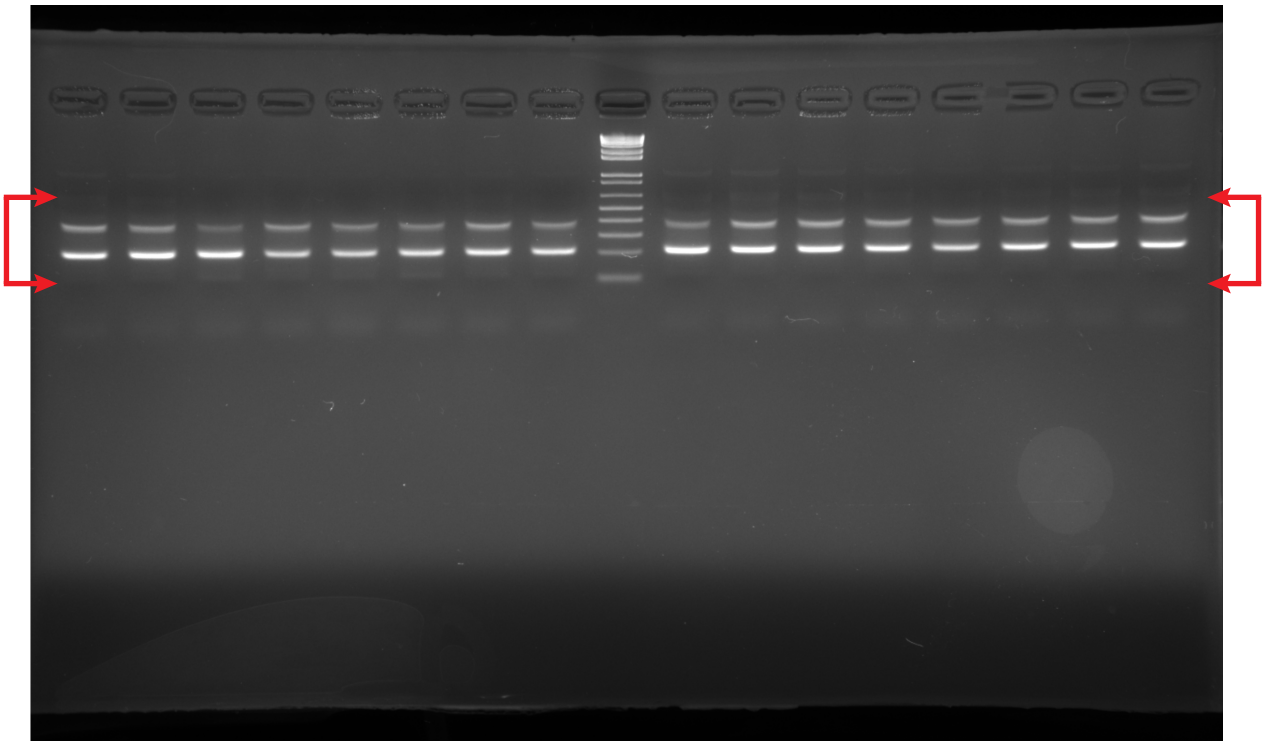
Right Bottom panels

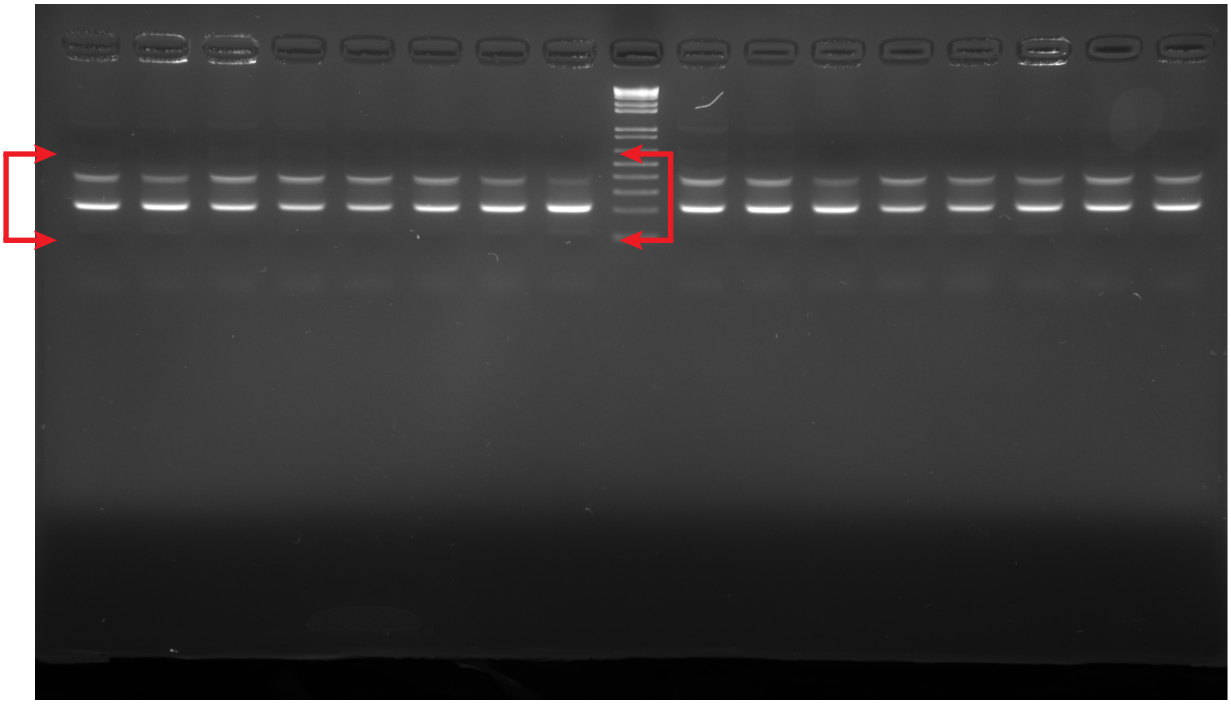






Full unedited gels for Supplemental Figure 3





Full unedited gels for Supplemental Figure 5