Supplemental information

Premature transcription termination at the expanded GAA repeats and aberrant alternative polyadenylation contributes to the *Frataxin* transcriptional deficit in Friedreich's ataxia

Yanjie Li, Jixue Li, Jun Wang, Siyuan Zhang, Keith Giles, Thazha P. Prakash, Frank Rigo, Jill S. Napierala, Marek Napierala

Table S1. Reagents and resources used in this study.

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
RNA polymerase II CTD repeat YSPTSPS (phospho S5) antibody	Abcam	Cat# ab5131
RNA polymerase II CTD repeat YSPTSPS antibody [8WG16]	Abcam	Cat# ab817
Anti-RNA polymerase II CTD repeat YSPTSPS (phospho S2) antibody	Abcam	Cat# ab5095
FXN antibody	Proteintech	Cat# 14147-1-AP
ACTIN monoclonal antibody	Santa Cruz	Cat# SC-47778
HRP-conjugated rabbit anti-mouse immunoglobulin	GE Healthcare life science	Cat# NA934V
HRP-conjugated donkey anti-rabbit immunoglobulin	GE Healthcare life science	Cat# NA931V
Chemicals, Peptides, and Recombinant Proteins		
Diethyl pyrocarbonate	Sigma-Aldrich	Cat# D5758
Sodium chloride	Sigma-Aldrich	Cat# S9888
Potassium chloride	Fisher Scientific	Cat# P330-500
Magnesium chloride	Acros Organics	Cat# 223211000
Calcium chloride	Sigma-Aldrich	Cat# C1016
Magnesium acetate	Sigma-Aldrich	Cat# M5661
Ammonium acetate	Sigma-Aldrich	Cat# A1542
Sodium acetate	Sigma-Aldrich	Cat# S2889
EDTA	Sigma-Aldrich	Cat# E9884
EGTA	Sigma-Aldrich	Cat# E3889
Sodium hydroxide	Fisher Scientific	Cat# S318-1
Triton X-100	Fisher Scientific	Cat# BP151
Nonidet P40	Sigma-Aldrich	Cat# 11332473001
Tween-20	Sigma-Aldrich	Cat# P9416
Tris base	Fisher Scientific	Cat# BP152-1
Hydrochloric acid	Fisher Scientific	Cat# A144S-500
DTT	Sigma-Aldrich	Cat# D0632
Glycerol	Sigma-Aldrich	Cat# G5516
SYBR Gold nucleic acid gel stain	ThermoFisher Scientific	Cat# S11494
Betaine	Sigma-Aldrich	Cat# B0300
Trizol	ThermoFisher Scientific	Cat# 115596018
Trizol LS	ThermoFisher Scientific	Cat# 10296028
Sarkosyl	Sigma-Aldrich	Cat# L5125
GlycoBlue	ThermoFisher Scientific	Cat# AM9515
Sucrose	Sigma-Aldrich	Cat# S0389
RNase inhibitor	ThermoFisher Scientific	Cat# AM2696
Protease Inhibitor Cocktail	Millipore-Sigma	Cat# P8340
Passive Lysis Buffer	Promega	Cat# E1941

Protein Assay Dye Reagent Concentrate	Bio-Rad	Cat# 500-0006
Chloroform	ThermoFisher Scientific	Cat# C606-4
Ethanol, absolute	ThermoFisher Scientific	Cat# BP2818500
Phenol/Chloroform/Isoamyl Alcohol (25:24:1 Mixture, pH 6.7/8.0)	ThermoFisher Scientific	Cat# BP1752-400
Superscript III reverse transcriptase	ThermoFisher Scientific	Cat# 18080044
T4 polynucleotide kinase (PNK)	New England Biolabs	Cat# M0201
RNA 5' Pyrophosphohydrolase (RppH)	New England Biolabs	Cat# M0356S
T4 RNA ligase I	New England Biolabs	Cat# B0300
Phusion polymerase	New England Biolabs	Cat# M0530
Lipofectamine 2000	ThermoFisher Scientific	Cat# 11668019
dNTP mix, 12.5 mM each	Roche	Cat# 03622614001
Biotin-11-CTP	PerkinElmer	Cat# NEL542001EA
ATP, 10 mM	Roche	Cat# 11277057001
GTP, 10 mM	Roche	Cat# 11277057001
UTP, 10 mM	Roche	Cat# 11277057001
25-bp DNA ladders	ThermoFisher Scientific	Cat# 10597011
Stemolecule™ Y27632	Stemgent	Cat# 04-0012
Triptolide (TRP)	Sigma-Aldrich	Cat# T3652-1MG
Flavopiridol Hydrochloride hydrate (FP)	Sigma-Aldrich	Cat# F3055-5MG
Actinomycin D	Sigma-Aldrich	Cat# A1410
Collagenase IV	ThermoFisher Scientific	Cat# 17104019
Dispase	STEMCELL Technologies	Cat# 07923
Accutase	STEMCELL Technologies	Cat# 07920
Trypsin EDTA, 0.25%	ThermoFisher Scientific	Cat# 25200056
Gelatin porcine skin	SIGMA	Cat# G1890
DMEM/F12	ThermoFisher Scientific	Cat# 11320082
Fetal Bovine Serum	ThermoFisher Scientific	Cat# SH30396.03
Nonessential amino acids	ThermoFisher Scientific	Cat# 11140
DMEM, high glucose	ThermoFisher Scientific	Cat# 11965118
Vitronectin, truncated human recombinant (VTN-N)	ThermoFisher Scientific	Cat# A14700
Corning Matrigel hESC	ThermoFisher Scientific	Cat# 354277
Clone R	STEMCELL Technologies	Cat# 05888
Essential 8™ Medium	ThermoFisher Scientific	Cat# A1517001
Critical Commercial Assays		
CytoTune™-iPS 2.0 Sendai Reprogramming Kit	ThermoFisher Scientific	Cat# A16517
TURBO DNA-free kit	ThermoFisher Scientific	Cat# AM1907
MTeSR1 Complete Kit, 500 mL kit	STEMCELL Technologies	Cat# 85850
Human Stem Cell Nucleofector® Kit	Lonza	Cat# VPH-5022
Power SYBR™ Green RNA-to-CT™ 1-Step Kit	ThermoFisher Scientific	Cat# 4389986
RNeasy Mini Kit	Qiagen	Cat# 74104

TURBO DNA-free kit	ThermoFisher Scientific	Cat# AM1907
FirstChoice® RLM-RACE Kit	ThermoFisher Scientific	Cat# AM1700M
TOPO TA/Sequencing PCR4 kit	ThermoFisher Scientific	Cat# 450030
MACE-seq kit	GenXPro GmbH	Cat# 160424.1
QIAquick Gel Extraction Kit	Qiagen	Cat# 28704
Deposited Data		
PRO-seq data	Available upon request	
ChIP-seq data	Available upon request	
MACE-seq data	Available upon request	
RNA-seq data	Available upon request	
Experimental Models: Cell Lines		
Human fibroblast (control and FRDA)	Napierala laboratory	(Li et al., 2016)
Human iPSC (control and FRDA)	Napierala laboratory	(Polak et al., 2012)
Human iPSC (GAA repeat homo-deleted)	This paper	
Drosophila S2 Cells	ThermoFisher Scientific	Cat# R69007
One Shot® Mach1™-T1R competent cells	ThermoFisher Scientific	Cat# C862003
Oligonucleotides		
Oligonucleotides for PRO-seq	This paper	Table S3
Oligonucleotides for 3-RACE	This paper	Table S3
Oligonucleotides for RT-qPCR	This paper	Table S3
Software and Algorithms		
SnapGene	GSL Biotech LLC	
GraphPad Prism v.6	GraphPad Software	
Partek Flow	Partek	
Integrative Genomics Viewer (IGV)	Broad Institute	
Alternative Splice Site Predictor (ASSP)	wangcomputing.com	
mfold	UNAFold	
Other		
P-30 column, RNase-free	Bio-Rad	Cat# 732-6250
Dynabeads™ Protein A for Immunoprecipitation	ThermoFisher Scientific	Cat# 100-02D
Dynabeads® M-280 Streptavidin	ThermoFisher Scientific	Cat# 11205D
NuPAGE™ 4 to 12%, Bis-Tris	ThermoFisher Scientific	Cat# NP0322BOX
Native PAGE gel 8% TBE	ThermoFisher Scientific	EC6215BOX

Table S2. Cell lines used in this study. FRDA iPSC and fibroblast cells are part of the FRDA Cell Line Repository (FACLR) (Li et al., 2016). FRDA 1 and CTRL 1 iPSCs were characterized in (Li et al., 2015), while FRDA 2, 3 and edited lines are described in detail in FACLR or in Supplemental Figure S7. Sizes of the GAA repeats in FRDA iPSCs and fibroblasts were determined as we described in (Polak et al., 2016).

Cell line	***GAA repeat length (GAA1, GAA2)***
iPSC	
FRDA 1	600, 600
FRDA 2	590, 830
FRDA 3	750, 750
CTRL 1	unaffected
FRDA 2 ΔGAA 3B12	corrected
FRDA 2 ΔGAA 3C7	corrected
FRDA 3 ΔGAA 2B1	corrected
FRDA 3 ΔGAA 2D3	corrected
Primary Fibroblasts	
F1	495, 505
F2	490, 680
F3	916, 1382
F4	630, 806
F5	816, 1410
F6	870, 1470
F7	404, 920
F8	526, 826
F9	211,1428
F10	468, 807
F11	190, 500
F12	185, 1130
F13	570, 1200
F14	422, 520
C1	unaffected
C2	unaffected
C3	unaffected
C4	unaffected
C5	unaffected
C6	unaffected
C7	unaffected
C8	unaffected
C9	unaffected
C10	unaffected
C11	unaffected
C12	unaffected
C13	unaffected

Table S3. Identification of the polyadenylation sites in FRDA and control iPSCs by MACE-seq. The most common PA is indicated in green. An alternative PA site identified only in FRDA iPSCs located immediately upstream of the expanded GAAs is highlighted in blue. A total of 24 and 57 curated reads were assigned in FRDA and control cells, respectively. * A single read in intron 1 of *FXN* was identified in CTRL iPSCs. This site however, is located downstream of the expanded GAAs. The bioinformatic pipeline used to define the APA sites is described in STAR★Methods.

FRDA 1 iPSC

Locus	Start of APA	End of PA	Chromosome	Feature	Strand	# of reads
FXN	**71652176**	**71652184**	**chr9**	**Intron**	**+**	**2**
FXN	71687982	71687987	chr9	UTR3	+	2
FXN	**71688012**	**71688032**	**chr9**	**UTR3**	**+**	**13**
FXN	71689255	71689256	chr9	UTR3	+	1
FXN	71689690	71689691	chr9	UTR3	+	1
FXN	71689804	71689809	chr9	UTR3	+	2
FXN	71693989	71693994	chr9	UTR3	+	3

Control 1 iPSC

Locus	Start of APA	End of PA	Chromosome	Feature	Strand	# of reads
FXN	71652417	71652418	chr9	Intron*	+	1
FXN	71687717	71687759	chr9	UTR3	+	5
FXN	71687973	71688006	chr9	UTR3	+	5
FXN	**71688012**	**71688030**	**chr9**	**UTR3**	**+**	**30**
FXN	71688133	71688141	chr9	UTR3	+	3
FXN	71688397	71688398	chr9	UTR3	+	1
FXN	71689129	71689130	chr9	UTR3	+	1
FXN	71689255	71689260	chr9	UTR3	+	2
FXN	71689694	71689695	chr9	UTR3	+	1
FXN	71689807	71689813	chr9	UTR3	+	3
FXN	71691494	71691495	chr9	UTR3	+	1
FXN	71693955	71693956	chr9	UTR3	+	1
FXN	71693988	71693999	chr9	UTR3	+	3

Table S4. Oligonucleotides used in this study.

Oligo	Sequence
3' RACE and PRO-seq oligonucleotides	
3' RACE adaptor	5' GCATCGTACGCGTACGTGTTTTTTTTTTTTNN
3' RACE primer	5' GCATCGTACGCGTACGTGTT
VRA3 adaptor	5' GAUCGUCGGACUGUAGAACUCUGAAC-/inverted dT/ The 5′ end is phosphorylated and the 3′ end is protected by an inverted dT
VRA5 adaptor	5' CCUUGGCACCCGAGAAUUCCA The 5′ end is not phosphorylated
RP1	5' AATGATACGGCGACCACCGAGATCTACACGTTCAGA GTTCTACAGTCCGA
RPI-1	5' CAAGCAGAAGACGGCATACGAGATCGTGATGTGACT GGAGTTCCTTGGCACCCGAGAATTCCA
RPI-2	5' CAAGCAGAAGACGGCATACGAGATACATCGGTGACT GGAGTTCCTTGGCACCCGAGAATTCCA
qRT-PCR primers	
FXN_F	5' CAGAGGAAACGCTGGACTCT
FXN_R	5' AGCCAGATTTGCTTGTTTGG
GAPDH_F	5' GAAGGTGAAGGTCGGAGTC
GAPDH_R	5' GAAGATGGTGATGGGATTTC
ACTIN_F	5' GCCAACCGCGAGAAGATGACC
ACTIN_R	5' CTCCTTAATGTCACGCACGATT
FXN-ett_F	5' GCACCGACATCGATGCGACC
FXN-ett_R	5' CACAGCCATTCTTTGGGTTTC
MYC_F	5' CCTGGTGCTCCATGAGGAGAC
MYC_R	5' CAGACTCTGACCTTTTGCCAGG
GAA PCR and CRISPR/Cas9 editing primers	
GAA_F	5' GGAGGGAACCGTCTGGGCAAAGG
GAA_R	5' CAATCCAGGACAGTCAGGGCTTT
UP_jct_F	5' AGATGAAAGAGGCAGGCCAC
UP_jct_R	5' GTGGCCTGCCTCTTTCATCT
DW_jct_F	5' CAACCTCTGTCACCAGCTGG
DW_jct_R	5' CCCACCACCATGCTGAGCTA
UP_DW_jct_F	5' GGTACAAAGAAATATCTGACCC
UP_DW_jct_R	5' CTGATTACATGTTGGAATGATA

Figure S1. Sanger sequencing of the 3' RACE PCR product representing *FXN* mRNA isoform type I. The most common, canonical form of the *FXN* mRNA was amplified by 3' RACE, excised from an agarose gel, cloned and sequenced. Locations of all exons, RACE primers and vector flanking sequences are indicated.

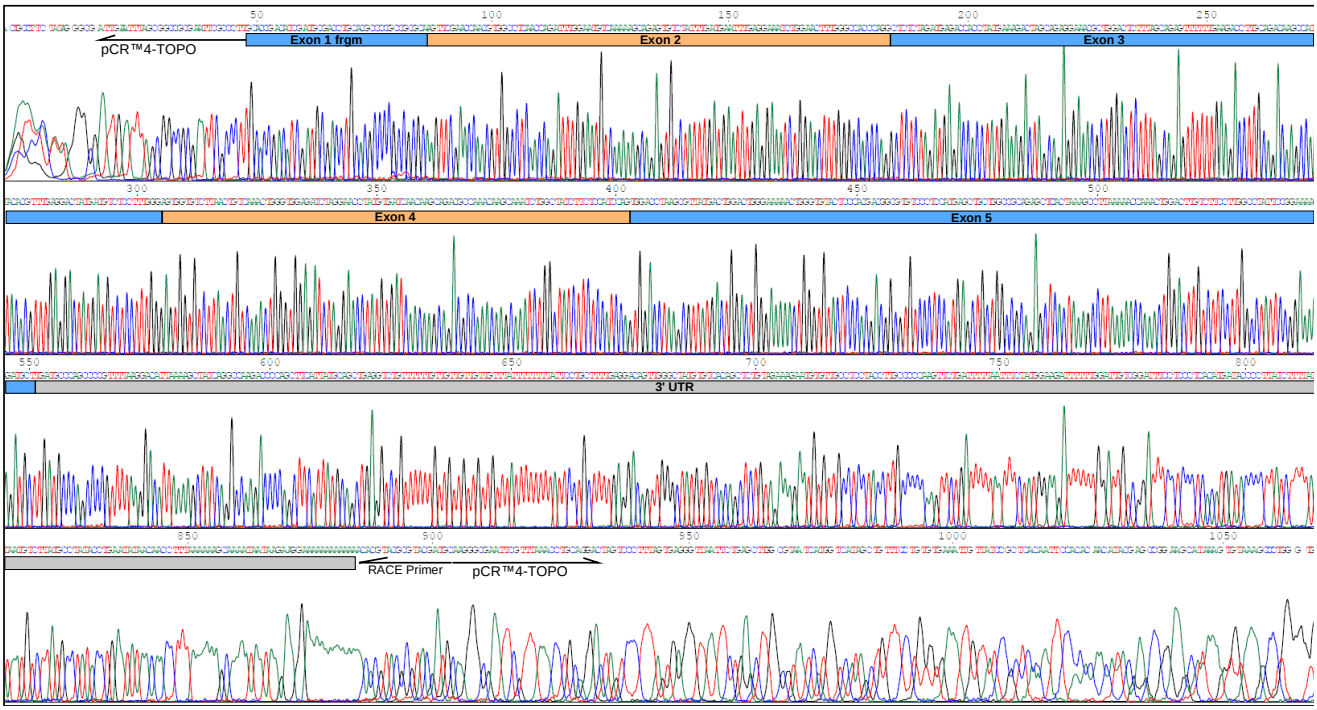


Figure S2. Sanger sequencing of the 3' RACE PCR product representing isoform type IV – *FXN-ett* RNA. The *FXN-ett* RNA is the predominant transcript detected using 3' RACE in FRDA cells. Exon 1, intron 1, location of the 683 nt splice fragment are indicated along with RACE primers and vector flanking sequences.

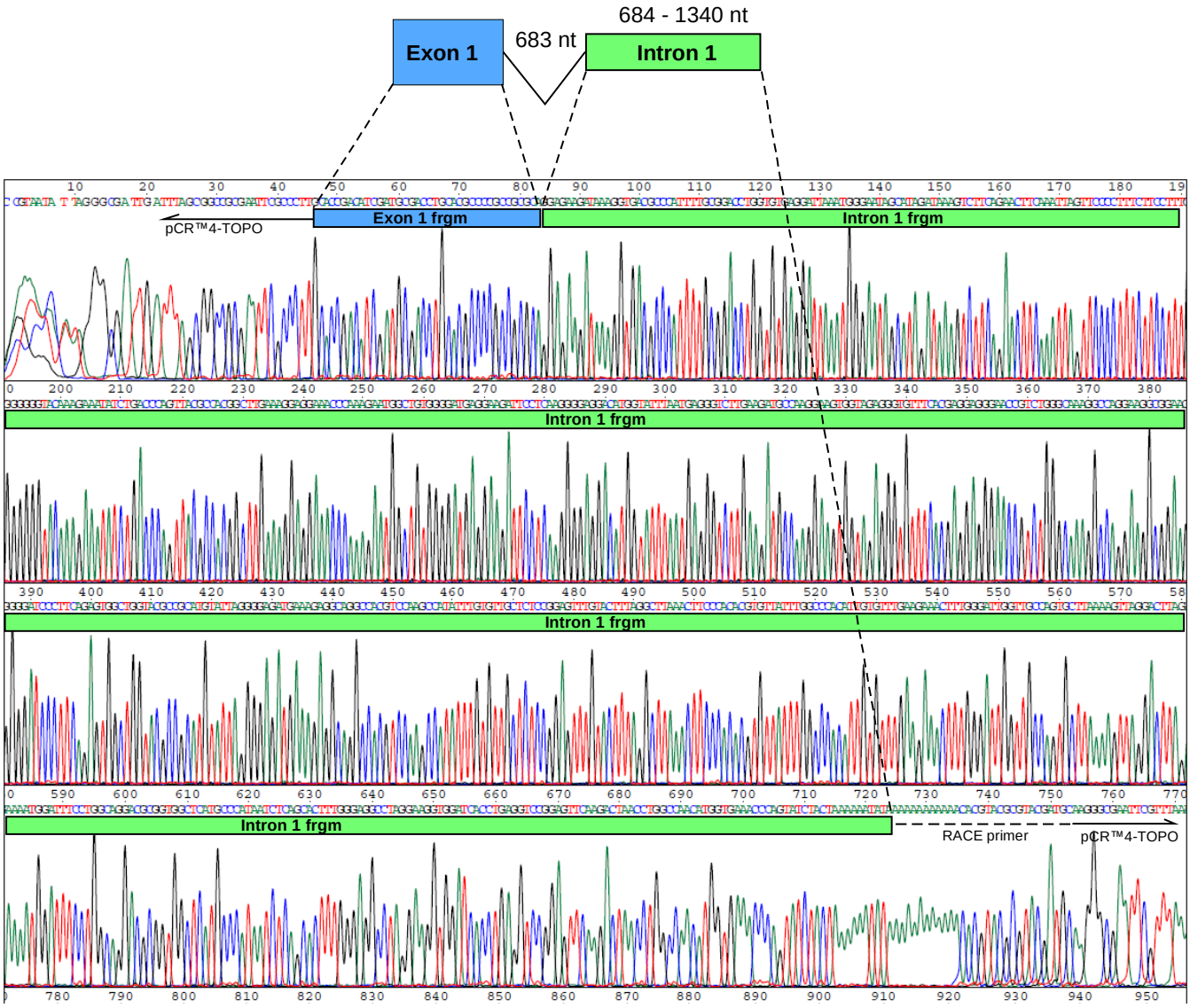


Figure S3. Figure S3. Mapping of RNA-seq reads to intron 1 of *FXN* in FRDA 1 and CTRL 1 iPSCs. Two reads corresponding to the spliced *FXN-ett* transcript are indicated by the arrow heads. Their 5' end sequence (indicated) starts at position 848 (A of the ATG START codon is designated as 1). No RNA-seq reads corresponding to the spliced *FXN-ett* were identified in CTRL 1 cells. The vertical black line indicates the position of the splice site.

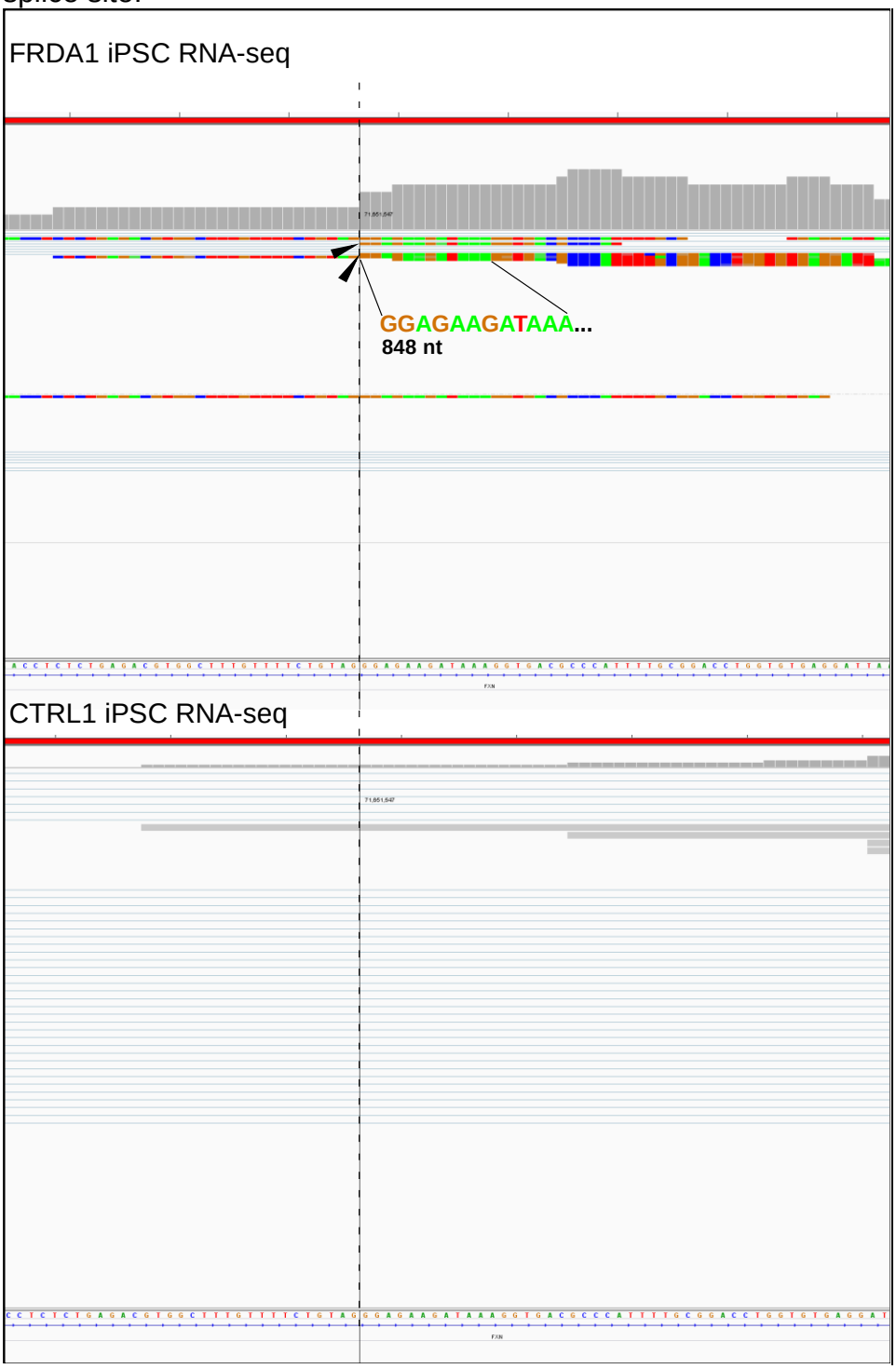


Figure S4. Figure S4. Prediction of the alternative splice sites in the proximal fragment of *FXN* intron 1 using Alternative Splice Site Predictor (ASSP) (Wang and Marin, 2006). The sequence of the entire exon 1 (starting at position 1 ATG) through intron 1 to the beginning of the GAA tract was used in the analyses. A. ASSP-predicted splice donor and acceptor sites. Constitutive donor site (D1, highlighted in yellow) and four putative acceptor splice sites are indicated as A1 – A4. Scores of the preprocessing models reflecting splice site strength. Constitutive acceptor A3 highlighted in red has the highest score reflecting the most probable acceptor in a given region. A3 is the acceptor site utilized in *FXN-ett* RNA. B. Location of the predicted splice donor and acceptor sites in the *FXN* mRNA sequence. The sequence of exon 1 is underlined. The ATG start codon is indicated in green font.

A

	Position (bp)	Putative splice site	Sequence	Score*
D1	165	Constitutive donor	CCGCCGCGCAgtaagtatcc	13.818
A1	302	Constitutive acceptor	tccttctcagGGCGGCCCGG	9.749
A2	654	Constitutive acceptor	tattttcaagGCTGGATTTT	3.737
A3	849	Constitutive acceptor	ttttctgtagGGAGAAGATA	11.260
A4	1365	Constitutive acceptor	ttcctggcagGACGCGGTGG	6.003

B

Atgtggactctcgggcgccgcgcagtagccggcctcctggcgtcacccagcccagcccaggccc
agaccctcacccgggtcccgcggccggcagagttggccccactctgcggccgccgtggcctgcg
caccgacatcgatgcgacctgcacgcc**ccgccgcgca**gtaagtatccgcgccgggaacagccgc
gggccgcacgccgcgggccgcacgccgcacgcctgcgcagggaggcgccgcgcacgccggggtc
gctccgggtacgcgcgctggactagctcaccccgctccttctcag**ggcggcccgg**cggaagcgg
ccttgcaactcccttctctggttctcccggttgcatttacactggcttctgctttccgaaggaa
aaggggacattttgtcctgcggtgcgactgcgggtcaaggcacgggcgaaggcagggcaggctg
gtggaggggaccggttccgaggggtgtgcggctgtctccatgcttgtcacttctctgcgataac
ttgtttcagtaatattaatagatggtatctgctagtatatacatacacataatgtgtgtgtctg
tgtgtatctgtatatagcgtgtgtgttgtgtgtgtgtttgcgcgcacgggcgcgcgcacacc
taatattttcaag**gctggatttt**tttgaacgaaatgctttcctggaacgaggtgaaactttcag
agctgcagaatagctagagcagcaggggccctggcttttggaaactgacccgacctttattcca
gattctgccccactccgcagagctgtgtgaccttgggggattcccctaacctctctgagacgtg
gctttg**ttttctgtagggagaagata**aaggtgacgcccattttgcggacctggtgtgaggatta
aatgggaataacatagataaagtcttcagaacttcaaattagttccccttttcttcctttggggg
gtacaaagaaatatctgacccagttacgccacggcttgaaaggaggaaacccaaagaatggctg
tggggatgaggaagattcctcaaggggaggacatggtatttaatgagggtcttgaagatgccaa
ggaagtggtagagggtgtttcacgaggagggaaccgtctgggcaaaggccaggaaggcggaagg
ggatcccttcagagtggctggtacgccgcatgtattaggggagatgaaagaggcaggccacgtc
caagccatatttgtgttgctctccggagtttgtactttaggcttgaacttcccacacgtgttat
ttggcccacattgtgtttgaagaaactttgggattggttgccagtgcttaaaagttaggactta
gaaaatggat**ttcctggcag gacgcggtgg**ctcatgcccataatctcagcactttgggaggcct
aggaaggtggatcacctgaggtccggagttcaagactaacctggccaacatggtgaaacccagt
atctactaaaaaatacaaaaaaaaaa

Figure S5. Figure S5. RNA structure prediction using a *FXN* pre-mRNA fragment. A RNA structure model was generated using mfold (Markham and Zuker, 2008). The sequence of the 5' UTR, entire exon 1 (orange) with START codon (green) through intron 1 to the beginning of the GAA tract was used for modeling. Location of the *FXN-ett* RNA splice acceptor site in a hairpin loop is indicated in blue and magnified.

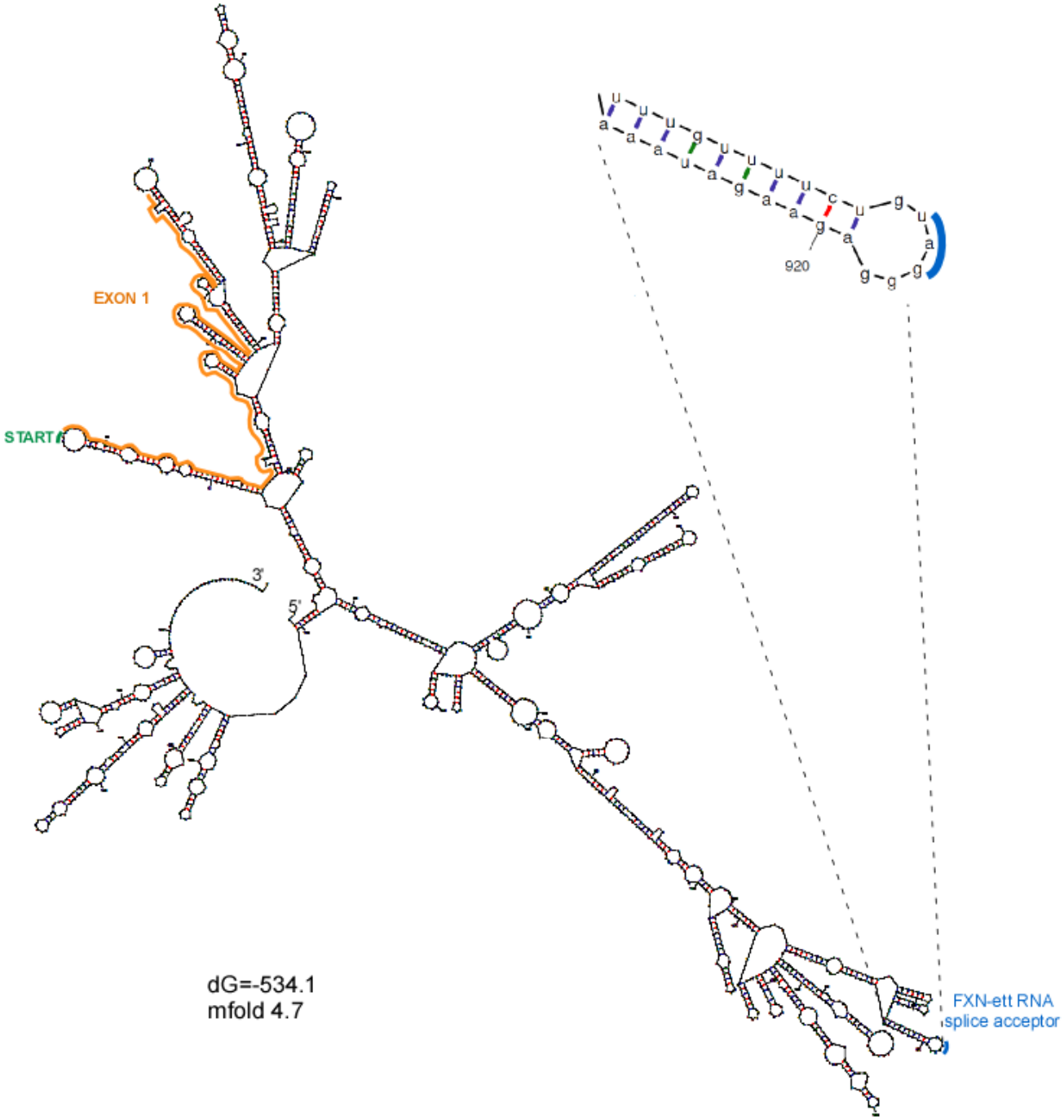


Figure S6. Figure S6. Half-life of *FXN-ett* RNA. The half-lives of *FXN-ett* RNA and *Myc* mRNA were determined in two FRDA fibroblast lines (F4 gray circles and F7 black squares) using an actinomycin D approach (described in STAR★Methods). Error bars correspond to technical replicates.

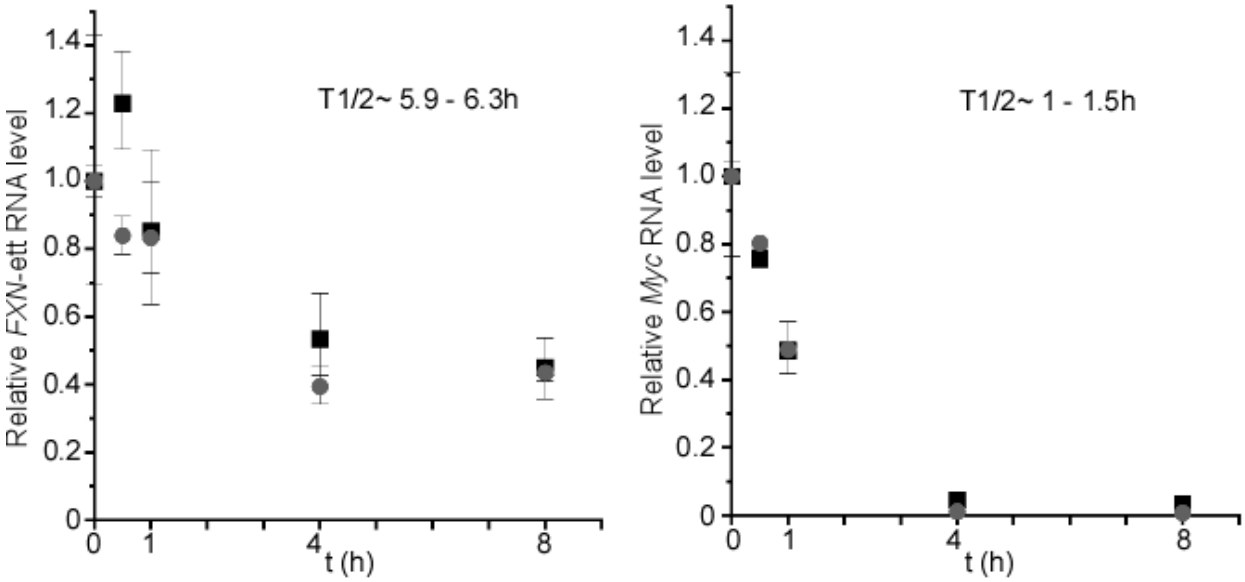
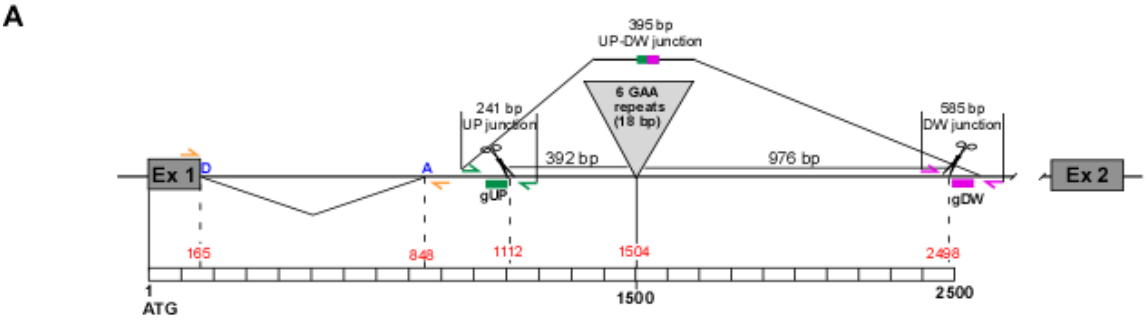
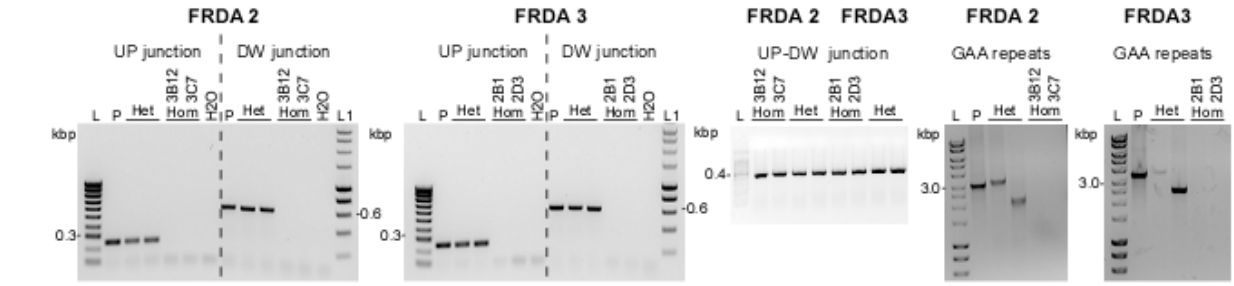
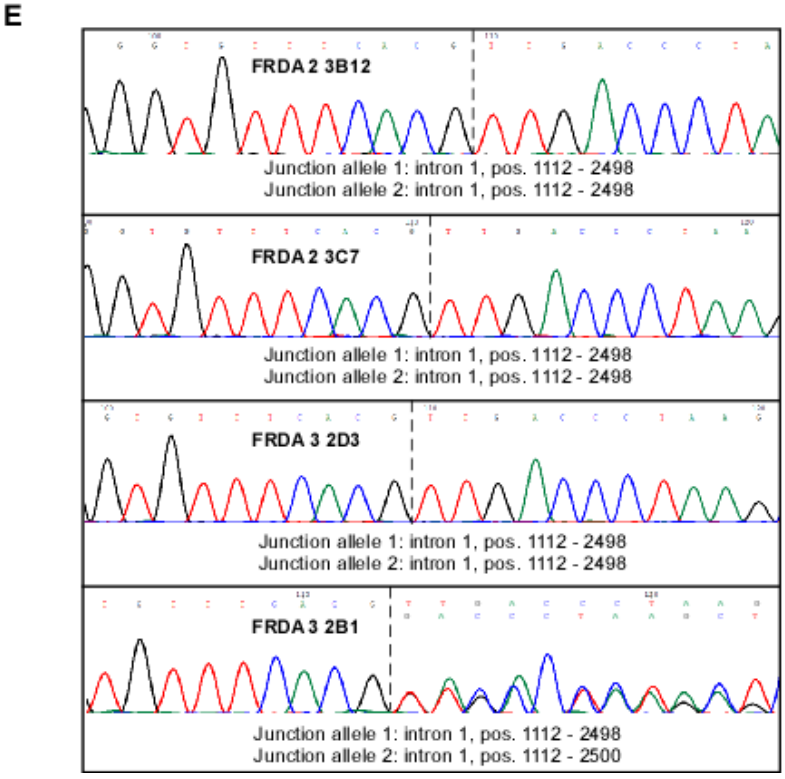


Figure S7. Excision of the expanded GAA tract from FRDA iPSCs using CRISPR/Cas9. A. Schematic illustrating localization of gRNAs in the *FXN* locus (drawn to scale). A guide upstream of the GAAs (gUP) is shown in green. The predicted Cas9 cut site is located at position 1112 bp (A of the ATG start is designated as 1). The length of the PCR product for the UP junction is indicated. A guide downstream of the GAAs (gDW) is indicated in purple together with amplicon length of the DW junction PCR and Cas9 cut site at 2498 bp. The position of the cut site is calculated based on the presence of only 6 GAAs (18 bp). Length of the UP-DW junction PCR amplicon after excision of the expanded GAAs together with flanking sequences is 395 bp. Lengths of the flanking sequences removed together with expanded GAAs are indicated in black. In addition, positions of the *FXN-ett* splice donor (blue D) and acceptor (blue A) sites are indicated along with approximate locations of qRT-PCR primers used for *FXN-ett* quantification. Sequences of both gRNAs are given below and all primers used for PCR analyses are listed in Table S3. B. PCR analysis of the UP and DW junctions in parental (P) and homozygous (Hom) edited clones: for FRDA 2 line Hom 3B12 and 3C7 and for FRDA 3 line Hom 2B1 and 2D3. Heterozygous (Het) edited lines (expanded GAA tract excised from only one of the two FXN alleles) are shown for reference. C. PCR analysis of the UP-DW junction in both Het and Hom edited clones. D. determination the GAA tracts sizes in P, Het and Hom FRDA 2 and FRDA 3 cells. E. Sanger sequencing of the UP-DW junction PCR products. The exact sequence of the junction resulting from Cas9-mediated GAA tract excision is indicated for each cell line. Line FRDA 3 clone 2B1 contains alleles that differ by two base pairs.



gRNAs:
gUP: GTAGAGGGTGTTTCACGAGGAGG
gDW: GGGGAGCTTAGGGTCAATCCAGG





Supplemental References

Li, Y., Polak, U., Bhalla, A.D., Rozwadowska, N., Butler, J.S., Lynch, D.R., Dent, S.Y., and Napierala, M. (2015). Excision of Expanded GAA Repeats Alleviates the Molecular Phenotype of Friedreich's Ataxia. Mol Ther *23*, 1055-1065.

Li, Y., Polak, U., Clark, A.D., Bhalla, A.D., Chen, Y.Y., Li, J., Farmer, J., Seyer, L., Lynch, D., Butler, J.S.*, et al.* (2016). Establishment and Maintenance of Primary Fibroblast Repositories for Rare Diseases-Friedreich's Ataxia Example. Biopreserv Biobank *14*, 324-329.

Markham, N.R., and Zuker, M. (2008). UNAFold: software for nucleic acid folding and hybridization. Methods Mol Biol *453*, 3-31.

Polak, U., Li, Y., Butler, J.S., and Napierala, M. (2016). Alleviating GAA Repeat Induced Transcriptional Silencing of the Friedreich's Ataxia Gene During Somatic Cell Reprogramming. Stem Cells Dev *25*, 1788-1800.

Wang, M., and Marin, A. (2006). Characterization and prediction of alternative splice sites. Gene *366*, 219-227.