

Supplemental Information

Hepatocyte-intrinsic SMN deficiency drives metabolic dysfunction and liver steatosis in spinal muscular atrophy

Damien Meng Kiat Leow¹, Yang Kai Ng^{1, 2}, Loo Chien Wang², Hiromi W. L. Koh², Tianyun Zhao², Zi Jian Khong², Tommasco Tabaglio², Gunaseelan Narayanan³, Richard M. Giadone⁴, Radoslaw M. Sobota², Shi Yan Ng^{1, 2, 5}, Adrian Kee Keong Teo^{1, 2}, Simon H. Parson⁶, Lee L. Rubin⁴, Wei YiOng¹, Basil T. Darras⁷, Crystal J. J. Yeo^{*2,3, 5-9}

¹ Yong Loo Lin School of Medicine, National University of Singapore, Singapore 117597, Republic of Singapore

² Institute of Molecular and Cell Biology (IMCB), Agency for Science, Technology and Research (A*STAR), 61 Biopolis Drive, Proteos, Singapore 138673, Republic of Singapore

³ Duke-National University of Singapore Medical School, Singapore 169857, Republic of Singapore

⁴ Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge MA02138, USA

⁵ National Neuroscience Institute, Singapore 308433, Republic of Singapore.

⁶ Institute of Education in Healthcare and Medical Sciences, School of Medicine, Medical Sciences and Nutrition, University of Aberdeen, Aberdeen, AB51 7HA, Scotland

⁷ Department of Neurology, Boston Children's Hospital, Harvard Medical School, Boston, Massachusetts, USA.

⁸ Department of Neurology, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA.

⁹ Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore 308232, Republic of Singapore.

Contents

Supplemental Information	1
Hepatocyte-intrinsic SMN deficiency drives metabolic dysfunction and liver steatosis in spinal muscular atrophy	1
Supplemental Methods	3
CRISPR/Cas9 Genome Editing.....	3
RT-qPCR Primers and Primer Efficiency	5
Oil Red O Image Analysis.....	7
Proteomics	8
Confocal imaging of iHeps.....	10
Supplemental Figures	12
Supplemental Figure 1.....	12
Supplemental Figure 2.....	13
Supplemental Figure 3.....	14
Supplemental Figure 4.....	15
Supplemental Figure 5.....	16
Supplemental Figure 6.....	17
Supplemental Figure 7.....	18
Supplemental Figure 8.....	19
Supplemental Figure 9.....	20
Supplemental Figure 10.....	21
Supplemental Figure 11.....	22
Supplemental Figure 12.....	23
Supplemental Figure 13.....	24
Supplemental Figure 14.....	25
Supplemental Figure 15.....	26
References	27
Supplementary Table S4	28

Supplemental Methods

CRISPR/Cas9 Genome Editing

Electroporation of sgRNA

70% confluent culture was used for electroporation. Cells were dissociated with TrypLE solution(Gibco) to obtain single cell suspension, and viability determined using Trypan Blue staining. 37.5 pmole sgRNA (IDT), 1 µg WT Cas9 protein (TrueCutTM Cas9 Protein v2; Invitrogen) and 1.25 µlR buffer (Neon transfection kit, Invitrogen) was mixed and incubated at room temperature for 20minutes to form RNP complexes. 0.5 µg ssODN donor (IDT) was then added and RNP complexeswere electroporated into 200,000 cells using the Neon transfection system (Invitrogen) following manufacturer's instructions. sgRNA spacer sequence TACAGGGTTTAGACAAAAT was used. ssODN sequence TGTGAGCACCTCCTCTTTGATTTGTCTGAAACCCTGTAAGGAAAATAAAGG AAGTTAAAAAAATAGCTATATAG was used. Electroporated cells were seeded between 500 to 5,000 cells/cm². On Day 3 post-electroporation, genomic DNA was extracted and editing efficiency was checked in the cell pool through PCR and Sanger sequencing. Genomic DNA was isolated using QuickextractTM DNA extraction solution (QE09050, Lucigen). PCR was performedusing the Q5® high-fidelity PCR kit (M0491L, NEB), following manufacturer's instructions. Sanger sequencing was outsourced to Bio Basic Asia Pacific Pte Ltd.

Isolating clonal colonies

On Day 5 post-electroporation, clonal colonies were transferred individually to each well in a 24-well dish. Colonies were expanded for a week before genomic DNA was isolated using QuickextractTM DNA extraction solution (QE09050, Lucigen).

Genotyping of clones

Genomic DNA was used for PCR and Sanger sequencing the region of interest. PCR forward primer GGTCTCAAGTGATCCCCCTACC, reverse primer TCTGCTGGTCTGCCTACTAGTG and sequencing primer GCATGAGCCACTGCAAGAAAAC were used. Sanger sequencing data of clones were analyzed using online tool ICE Analysis v2.0 (<https://ice.synthego.com/#/>) to identify corrected clones.

Off-target analysis in clones

The top 10 off-target sites for the sgRNA (predicted by Benchling) were Sanger sequenced to check for off-target effects. Sequencing data were analyzed using online tools ICE Analysis v2.0 (<https://ice.synthego.com/#/>). The primers used for amplification and sequencing are listed in **Table S1**.

Cryopreservation

Cells were manually cut into clumps using a pipette tip. Clumps were then scraped using a cell scraper and transferred into cryovials. Cell clumps were short spun and supernatant removed. Cell clumps were resuspended in Cryostor® freezing medium (Stemcell Technologies). Cells were kept at -80°C overnight and transferred to liquid nitrogen for long-term storage.

Supplementary Table S1. Sequencing and PCR primers for off-target sites

Site No.	Sequence	PAM	Chromosome	Strand	Position	Sequencing Primer	PCR Forward Primer	PCR Reverse Primer
1	AGTCT GTGTA AAACC CTGTA	AGG	chr1	1	107759751	ACTCCAT AGAATTG TGAGAGG G	ACTATGTT TGGCACAG AGAACGCC	AGTTTCCA AGGCATCC TGGAAC
2	ACTTG GTCTG AAACC CTGTA	TAG	chr14	-1	66452231	TGTCTTAA GAGAACCC TTTGACTG	TTGCCTCT TACCAAGAA CCCCAG	TGCCCTGC TCTGCTTT CAAATC
3	ATTCA GTCTT AAACC CTGTA	GGG	chr4	-1	164954141	TCCTGGT GCAATCC TCACCTCA G	CACCCCAT CACAAACCA TTCTGC	CATGCTGC CAACCATG TGTAGG
4	AGCCT GTGTA AAACC CTGTA	GAG	chr2	-1	241204863	GTGTCCA GAGTCTA GGGCAGAG G	TTTCACAG CAGTCACT TTGGCC	AGGGAAC ACTCACGT GGGTATG
5	ATCTT GTCTA AAAAC CTGTA	TGG	chr3	1	99910195	GGGGCAA AATGTGT ATTCAAGA GC	GTCTGCAA CCCCTTTC CTGTTTC	AACATGC ATCTGCCA CTTCACC
6	TTTTG ACTAT AACCC TGTA	TGG	chrX	-1	117007482	AGCACAA ACAATTG ACACATT CC	TCATCACC ACAATGCA TGCAGG	CCTCTTGG TGCCTAGC AGAAAC
7	CATTT GGGTA AAACC CTGTA	TAG	chrX	1	12823237	TGACTTTT AACAGGT ACAGGTG	AGTGTCCA GCAAGCCT ACTCTG	CCCCAAA ATGTGTCA GTGTCC
8	ATTTT GTGTG AAGCC CTGTA	AGG	chr3	-1	27994872	TCTCTGA AACTTTA ATGTGAC TTG	ACTTGCAC TAGAGGA AGTCTAAA TTCC	TGACGAGT TAGTGGGT GCAGTG

9	ATTCT GGGTT AAACC CTGTA	TGG	chr6	-1	147690888	CAACAAA GTATGGT GGAAATT CAG	GGAGTTCA AGAAAAG CCTGGCC	AGCTGGAA AGGAAAC AAAGGCAC
10	CATTT GTCTC AAACC CTGTG	GAG	chr20	-1	52447631	CCTACTG CAGTCCT ACACACA G	AGTTCTTC CCATGTCT CCCAGC	GCTGGGAT TACAAGC ATGAGCC

RT-qPCR Primers and Primer Efficiency

Supplementary Table S2. Primers used for RT-qPCR.

S/N	Primers (<i>Homo Sapiens</i>)	Sequences (5' to 3')
1	<i>GAPDH</i>	Forward: GCAAATTCCATGGCACCGT Reverse: GCCCCACTTGATTTGGAGG
2	<i>SDHA</i>	Forward: GAGATGTGGTGTCTCGGTCCAT Reverse: GCTGTCTCTGAAATGCCAGGCA
3	<i>SDHB</i>	Forward: GCAGTCCATAGAACAGCGTGAG Reverse: TGTCTCCGTTCCACCAGTAGCT
4	<i>SDHC</i>	Forward: CTGTTGCTGAGACACGTTGGT Reverse: ACAGAGGACGGTTAACCTA
5	<i>MT-CO1</i>	Forward: CCGTCCTAACATCACAGCAGTCCTA Reverse: TGAGGTTGCGGTCTGTTAGTAGT
6	<i>MT-CO2</i>	Forward: CCGCCATCATCCTAGTCCTCAT Reverse: GATCGTTGACCTCGTCTGTTATGT
7	<i>ATP5A</i>	Forward: TTTGAATCCCTATGAAGCGTT Reverse: CCTTGGGTATTGCTTAATCGACC
8	<i>MT-ATP6</i>	Forward: TGATCATTCTATTCCCCCTCT Reverse: GTCATTGTTGGGTGGTGATT
9	<i>ALB</i>	Forward: GTTGCATGAGAAAACGCCAGT Reverse: GTCGCCTGTTCACCAAGGAT
10	<i>ASGR1</i>	Forward: GAGAGAGACGTTCAGCAACTTC Reverse: GGGACTCTAGCGACTTCATCTT
11	<i>ASGR2</i>	Forward: GCTTCAGCAACTTCTCCTC Reverse: TTCTCCAGCTTGGCTCCTA
12	<i>APOE</i>	Forward: TCGCTTTGGGATTACCT Reverse: CTCCTTCATGGTCTCGTC
13	<i>APOA1</i>	Forward: ATGTGTCCCAGTTGAAG Reverse: CCCTCTGTCTCCTTTCC
14	<i>F2</i>	Forward: TTGCTGCATGTCTGGAAGGTA Reverse: GGATGGGTAGTGGAGTTGATT
15	<i>SERPINA1</i>	Forward: CCAAGGCCGTGCATAAGG Reverse: GGCCCCAGCAGCTTCAG
16	<i>HP</i>	Forward: TTTTGCACTGGACTCAGG Reverse: GCGCAGTTGTAGTAGTT
17	<i>HNF4A</i>	Forward: CTCCATCAATGCGCTCCT Reverse: CTTCATGGACTCACACACATCT
18	<i>AFP</i>	Forward: AGTGAGGACAAACTATTGGCCT Reverse: ACACCAGGGTTACTGGAGTC
19	<i>OCT4</i>	Forward: AGAAGTGGGTGGAGGAAG Reverse: ACGAGGGTTCTGCTTT

20	<i>NANOG</i>	Forward: ATGCCTCACACGGAGACTGT Reverse: AGGGCTGTCCTGAATAAGCA
21	<i>SOX2</i>	Forward: GAAAGAAAGGGAGAGAAGTTGAG Reverse: GCAAACATGGAATCAGGATCAA
22	<i>CPT1A</i>	Forward: CCAGACGAAGAACGTGGTCA Reverse: ATCTGCCGTGCTCAGTGAA
23	<i>CPT2</i>	Forward: GTAGCACTGCCGCATTCAAG Reverse: GCCATGGTACTTGGAGCACT
24	<i>ACSL1</i>	Forward: GACGAGCCCTTGGTGTATTT Reverse: TTCATAGGGTTGGTCTGGTTTC
25	<i>ACADI</i>	Forward: TGTGCCTATTGTGTAACAGAACCT Reverse: TCCATTGGTGATCCACATCTCTGAC
26	<i>ACOT1</i>	Forward: GGTGACCAAAGATGGCTATG Reverse: TGACCTACCAGGAACAGGAA
27	<i>HADHA</i>	Forward: ACTGCTGTCCTCTTCAGCTCAA Reverse: ACTGACTGAGCGAGGCATGA
28	<i>SCD1</i>	Forward: CCGGGAGAATATCCTGGTT Reverse: CGGGTACTCAACTGGCAGAGT
29	<i>SREBP1</i>	Forward: ACAGTGACTTCCCTGGCCTAT Reverse: GCATGGACGGGTACATCTTCAA
30	<i>HMGCS1</i>	Forward: TGTACACATCTTCAGTATATGGTTCCC Reverse: AAGAAAACACTCCAATTCTCTCCCT
31	<i>PCK2</i>	Forward: AGCCTCTCCACCTGGTGT Reverse: AATCGAGAGTTGGGATGTGC
32	<i>G6Pase</i>	Forward: GGGAAAGATAAAGCCGACCTAC Reverse: CAGCAAGGTAGATTGTGACAG
33	<i>FMO1</i>	Forward: ACCTGGCGAAAAGGTGT Reverse: CATGTTCTGAAAGCGTGTCA
34	<i>FMO3</i>	Forward: TTCCCACAGTTGACCTCC Reverse: CCATTGCGCTTTCTCC
35	<i>ACTB</i>	Forward: AGATGAGATTGGCATGGCTTTA Reverse: GGACTTCCTGTAACAACGCATC
36	<i>HPRT</i>	Forward: TGCTGAGGATTGGAAAGGG Reverse: ACAGAGGGCTACAATGTGATG

Supplementary Table S3. Evaluation of RT-qPCR primer efficiencies.

S/N	Gene Name	PCR Primer Efficiency (CT range)	R ²	Intra-assay Variation (% CV)	Range of CT values across all iHeps/ positive control/ negative control	
					Min.	Max.
1	<i>GAPDH</i>	91% (20.2-35.0)	0.99	0.8	15.2	30.6
2	<i>BETA-ACTIN</i>	99% (19.0-36.5)	0.99	0.6	18.0	22.4
3	<i>HPRT</i>	95% (19.9-34.9)	0.98	0.9	19.4	26.9
4	<i>SDHA</i>	92% (25.5-37.0)	0.98	0.4	19.8	35.2
5	<i>SDHB</i>	91% (24.6-36.3)	0.98	1.0	19.2	33.2
6	<i>SDHC</i>	109% (25.5-35.8)	0.98	0.6	19.2	36.7
7	<i>MT-ATP6</i>	93% (17.9-35.8)	0.99	0.5	11.3	23.6
8	<i>ATP5A</i>	98% (27.3-38.5)	0.98	0.7	20.6	31.5
9	<i>MT-CO1</i>	86% (15.8-35.1)	0.99	1.1	12.0	24.1
10	<i>MT-CO2</i>	87% (18.1-36.5)	0.99	0.4	12.6	24.5
11	<i>OCT4</i>	103% (21.1-35.2)	0.98	0.6	22.9	36.6

12	<i>SOX2</i>	110% (24.0-37.0)	0.98	0.2	27.4	39.6
13	<i>NANOG</i>	96% (24.5-35.7)	0.98	0.6	23.6	32.1
14	<i>ASGR1</i>	92% (24.5-35.9)	0.98	< 0.1	19.3	33.8
15	<i>ASGR2</i>	88% (24.4-35.1)	0.99	1.0	18.0	37.6
16	<i>ALB</i>	93% (23.9-34.3)	0.98	1.0	14.6	35.6
17	<i>HNF4A</i>	94% (27.2-34.7)	0.98	0.6	22.5	39.1
18	<i>HP</i>	92% (27.8-38.0)	0.99	0.8	21.4	37.9
19	<i>AFP</i>	87% (23.0-38.3)	0.99	0.4	12.8	32.6
20	<i>F2</i>	110% (25.7-35.4)	0.98	0.6	20.5	38.6
21	<i>SERPINA1</i>	105% (18.4-38.6)	0.98	0.8	17.1	36.1
22	<i>ACAD1</i>	101% (28.6-35.8)	0.98	0.4	22.5	35.3
23	<i>CPT1A</i>	86% (28.0-38.6)	0.99	0.4	20.4	36.6
24	<i>CPT2</i>	93% (19.7-37.0)	0.98	0.5	23.5	38.0
25	<i>HADHA</i>	110% (24.3-37.4)	0.98	0.6	18.9	32.4
26	<i>APOE</i>	93% (26.1-37.2)	0.98	0.8	20.5	28.7
27	<i>APOA1</i>	91% (22.0-37.1)	0.99	0.5	17.3	38.2
28	<i>ACOT1</i>	91% (26.1-35.8)	0.98	0.6	20.7	36.5
29	<i>PCK2</i>	102% (24.8-35.5)	0.98	0.6	19.6	35.7
30	<i>ACSL1</i>	97% (27.7-35.1)	0.98	0.4	22.2	31.6
31	<i>SCD1</i>	94% (25.5-37.6)	0.98	0.1	17.9	25.3
32	<i>SREBP1</i>	88% (20.6-37.0)	0.98	1.1	28.3	39.7
33	<i>G6PASE</i>	94% (24.1-36.0)	0.98	0.7	25.2	39.2
34	<i>FMO1</i>	98% (27.4-37.0)	0.98	0.8	24.3	38.5
35	<i>FMO3</i>	91% (24.7-37.0)	0.98	1.1	29.8	39.9
36	<i>HMGCS1</i>	100% (24.9-37.0)	0.98	1.8	20.9	24.7

The amplification efficiency for each primer threshold cycle (Ct) and the logarithm of the initial cDNA concentrations were plotted to calculate the slope (S) of each primer pair. Standard curves were generated from at least four dilution points for each primer pair. RT-qPCR reactions for each sample were run in duplicate, with standard deviations <0.85. Primer concentration (μM), standard deviation (SD), co-variance (CV), amplification efficiency (E) and correlation coefficient (R^2) * qPCR efficiency ($E = 10^{(-1/\text{slope})} - 1$) and correlation coefficient (R^2) were determined by standard curve by excel data.

Oil Red O Image Analysis

All histological analyses were done by a blinded researcher. Oil Red staining was measured using morphometry (ImageJ, version 2.7.0). Images taken with a microscope (three fields per sample) were randomly selected in ImageJ and transformed to black and white using "Image-type" submenu. Next, threshold function was applied using either the default or automatic settings to only choose the dark portion of the image (i.e., image area containing Oil Red staining). "Analyze- Analyze particles" submenu was used to calculate total Oil Red staining area. Cell counting per microscopic field was done by a blinded researcher based on haematoxylin stain. Total red intensity was normalised to cell numbers per microscopic field. The same image acquisition parameters and analysis methods were used for quantification of all samples.

Proteomics

Sample Preparation for Proteomics

Cell pellets were collected after 24 days of hepatocyte differentiation from human stem cells. Cell pellet samples were lysed using 8 M urea in 50 mM Tris for protein denaturation followed by probe sonication (Vibracell, Sonics and Materials, Inc.) for 3 cycles of 15 s pulse at 30% amplitude. Total protein amount in all samples were quantified by bicinchoninic acid assay (BCA) measurement and 50 µg of protein was taken for downstream experiment. Samples were reduced using a final concentration of 10 mM tris (2-carboxyethyl) phosphine (TCEP) and alkylated using a final concentration of 55 mM 2-chloroacetamide (CAA) at 25 °C for 30 min in the dark. This was followed by digestion with 1.5 µg of endoproteinase LysC for 4 h and 1.5 µg trypsin overnight at 25 °C. Digestion was terminated by adding 1% (v/v) final concentration of trifluoroacetic acid (TFA) to the samples, followed by desalting with Oasis HLB plate (1 cc/30 mg sorbent, Waters Corporation). The wells were activated with 100% (v/v) acetonitrile and equilibrated with 0.1% (v/v) formic acid in water. Peptides were then loaded onto the wells and washed once with 0.1% (v/v) formic acid in water. Bound peptides were eluted with 65% acetonitrile in 0.1% formic acid into a 2-ml deep well plate and transferred to 1.5-ml Eppendorf tubes later for drying.

Desalted peptides were resuspended with 100 mM triethylammonium bicarbonate (TEAB) and grouped as two sets to be labelled with isobaric tandem mass tags (TMT16-plex). The labelling was done for 2 h at 25 °C. Samples were then fractionated by high-pH reversed-phase at different concentrations of acetonitrile in 10 mM ammonium formate, pH (14%, 18%, 21%, 24%, 27%, 32% and 60%). Eluted peptides were dried by centrifugal evaporation for MS analysis.

Tandem Mass Spectrometry Analysis

Dried samples were resuspended in 10 µl of 2% (v/v) acetonitrile containing 0.06% (v/v) trifluoroacetic acid and 0.5% (v/v) acetic acid and transferred to an autosampler plate. Online chromatography was performed in an EASY-nLC 1000 (Thermo Fisher Scientific) liquid chromatography system using single-column setup and 0.1% formic acid in water and 0.1% formic acid in 99% acetonitrile as mobile phases. Fractions were injected and separated on a reversed-phase C18 analytical column (Easy-Spray, 75 µm inner diameter × 50 cm length, 2 µm particle size, Thermo Fisher Scientific) maintained at 50 °C and using a 2-27% (v/v) acetonitrile gradient over 45 min, followed by an increase to 55% over the next 15 min, and to 95% over 5 min. The final mixture was maintained on the column for 5 min to elute all remaining peptides. Total run duration for each sample was 70 min at a constant flow rate of 300 nl/min.

Data was acquired using an Orbitrap Fusion Eclipse mass spectrometer (Thermo Fisher Scientific) using data-dependent mode. Samples were ionized using 2.1 kV and 300 °C at the nanospray source. Positively-charged precursor signals (MS1) were detected using an Orbitrap analyzer set to 60,000 resolution, automatic gain control (AGC) target of 400,000 ions, and maximum injection time (IT) of 100 ms. Precursors with charges 2–7 and having the highest ion counts in each MS1 scan were further fragmented using higher-energy collision-induced dissociation (HCD) at 42% normalized collision energy. Fragment signals (MS2) were analysed by the ion trap analyzer at an AGC of 75,000 and maximum IT of 50 ms. Precursors used for MS2 scans were excluded for 90 s to avoid re-sampling of high abundance peptides. The MS1–MS2 cycles were repeated every 3 s until completion of the run.

Proteomics Data Analysis

Proteins were identified using Proteome Discoverer™ (v3.0, Thermo Fisher Scientific). Raw mass spectra were searched against human primary protein sequences retrieved from UniProt (11 June 2019). Carbamidomethylation on Cys and TMT16-plex were set as a fixed modification; deamidation of Asn and Gln, acetylation on protein N termini, and Met oxidation were set as dynamic modifications for the search. Trypsin/P was set as the digestion enzyme and was allowed up to three missed cleavage sites. Precursors and fragments were accepted if they had a mass error within 10 ppm and 0.06 Da, respectively. Peptides were matched to spectra at a false discovery rate (FDR) of 1% (strict) and 5% (relaxed) against the decoy database. Database search identified a total of 4,306 human protein groups within the search parameters used (see “Total protein ID” in **Supplementary Table S4**). Search results were exported and further processed using an in-house R script that utilizes *limma* for differential analysis, similar to a study by Ritchie *et al.*, 2015. For quantitative analysis, no imputation was performed for proteins that showed missing value in any of their TMT channels and they were removed from downstream processing. Median normalization was carried out across the samples to equalize the medians and eliminate potential differences arising from differing sample loading amounts (**Supplementary Fig. S13A**). Principal component analysis (PCA) was used to visualize the variability of the samples and determine outliers (**Supplementary Fig. S13B**). Normalized data was applied to *limma* R package for differential expression analysis using linear model. Statistical significance is defined at false discovery rate (FDR) < 0.05 and a fold-change requirement of equal to or greater than 1.5-fold difference is applied to differential proteins. For comparison sets where there were ≤ 200 proteins, it was not possible to adjust for multiple hypotheses and *p*-values < 0.01 were used for significant differences. Of the 6567 proteins in the cell pellet samples that showed significant differences in expression in at least one of the SMA

phenotypes (SMA0, SMA1, SMA2, or SMA3) compared to WT from the linear model analysis, we further investigated whether these proteins showed increasing or decreasing trends with known clinical severities. We fitted a linear regression on the expression of the proteins against the phenotypic groups as ordinal variables with increasing severity (from WT to SMA0) and tested the hypothesis if the slope is equal to zero, which would indicate that there is no correlation of the proteins with disease severity. At a stringent FDR of < 0.01 , the analysis revealed 343 differentially-expressed proteins exhibiting trends correlating with progressive severity of the disease (**Supplementary Table S4**).

For the iHeps cell pellet samples, proteins with significant differences were visualized for their interactions using STRING database (<https://string-db.org/>). Briefly, the significantly different genes were submitted and matched to the *Homo sapiens* database to correctly identify available proteins. Default settings were used except for the following: meaning of network edges was set to “evidence” and “textmining” was removed from active interaction sources to keep experimentally-validated edges only with confidence cut-off at 0.400. *k*-means clustering was then performed to separate the network into three distinct clusters (**Fig. 3A** and **Supplementary Fig. S9A and S12**). Proteins in individual clusters were analyzed for Gene Ontology (GO) biological processes term enrichment using the ShinyGO app hosted by South Dakota State University (<http://bioinformatics.sdstate.edu/go/>, v0.77). The top 20 pathways were shown, and pathways were selected by their $-\log_{10}(\text{FDR})$ scores and sorted by fold enrichment in decreasing order (**Supplementary Fig. S6 - 8**).

Following ShinyGO analysis, it was found that mitochondria-related processes were enriched for Cluster 1 (**Figure 3A, red**) proteins. Following up on this, proteins from the 343 differentially-expressed list were matched to only those that are associated with mitochondria based on the subcellular localization on the UniProt database (1,293 mitochondrial proteins out of 20,424 proteins in the proteome) and mitochondrial pathway using MitoCarta3.0 mitochondrial pathway inventory (<https://www.broadinstitute.org/mitocarta/>). ShinyGO enrichment analysis was performed again for this subset of 54 proteins. Proteins directly involved in mitochondrial complex V (ATP synthase) and lipid homeostasis (Lipid transport, triglyceride accumulation, fatty acid oxidation and cholesterol metabolism) respectively based on UnitProt database, were also analysed using STRING database.

Confocal imaging of iHeps

Immunocytochemistry was employed to visualize mitochondrial density and mitochondrial membrane potential. WT BJ iPSCs and SMA1 GM24468 iPSCs (Zero *SMN1* copy number, two

SMN2 copy number and age of biopsy at age of three) were (Coriell, New Jersey, United States) were seeded onto 13mm coverslips in respective 6-well plates prior to hepatocyte differentiation illustrated in **Figure 1**. After at least 24 days of differentiation, iHeps were washed thrice with 1X PBS before incubationwith 50nM of MitoTracker™ Green FM (Thermo Fisher Scientific) and tetramethylrhodamine methyl ester perchlorate (TMRM) (Sigma Aldrich) for 15min. Three 1X PBS washes were performed before addition of ProLong™ Gold Antifade Mountant with DAPI (Thermo Fisher Scientific) and visualized with Olympus FV3000 confocal microscope (Tokyo, Japan). Percentages of normalized TMRM intensity to MitoTracker™ (TMRM/MitoTracker™) were plotted using ImageJ v1.51 and GraphPad prism 9 software. The same image acquisition parameters and analysis methods were used for quantification of all samples similar to Stuhr et al.[1]

Supplemental Figures

Supplemental Figure 1

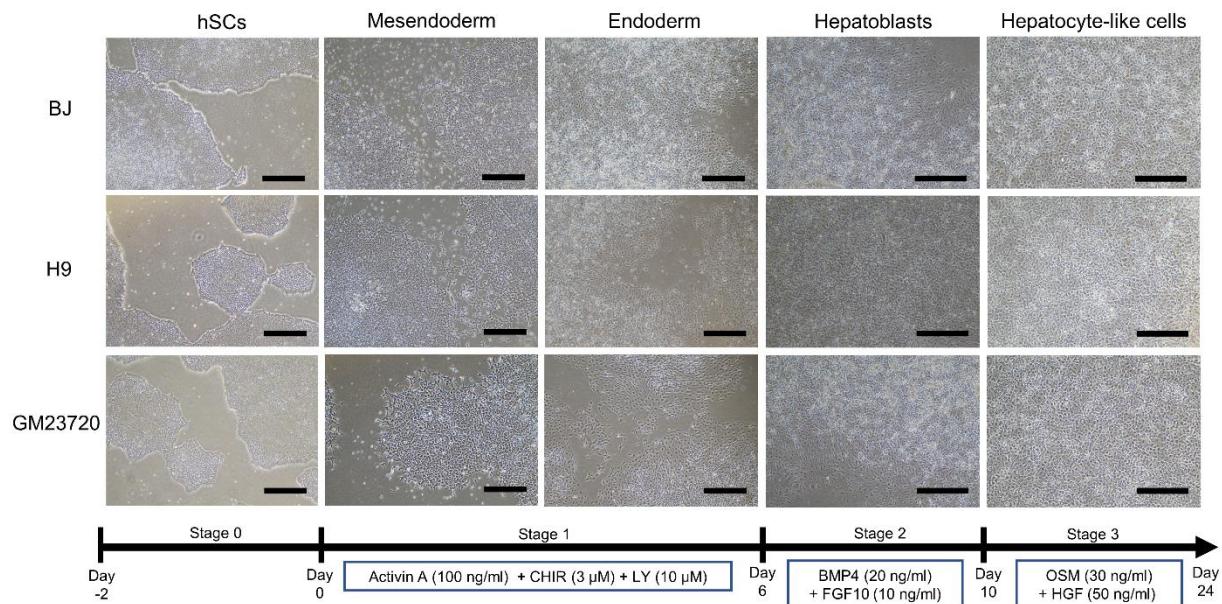


Figure S1. Differentiation of patient-derived stem cells to Day 24 iHeps. Morphological representation of WT stem cells through the iHep differentiation with schematic of differentiation over 24 days. Day 0: Stem cells; Day 2: Mesendoderm; Day 6: Definitive endoderm; Day 10: Hepatoblasts; Day 24: Hepatocyte-like cells (iHeps). Scale bar for hSCs, mesendoderm and endoderm: 500 μ m. Scale bar for hepatoblasts and hepatocyte-like cells: 250 μ m.

Supplemental Figure 2

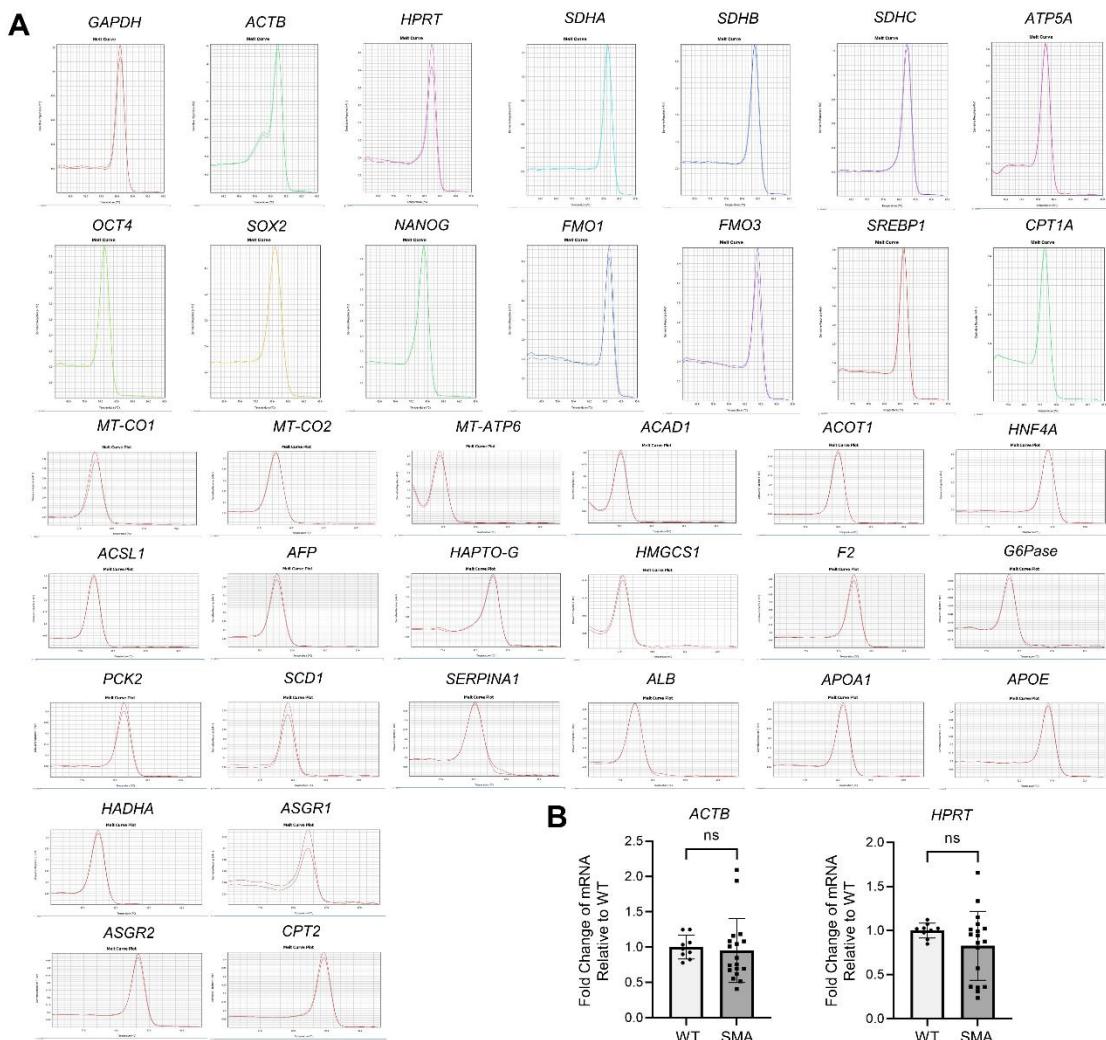


Figure S2. Melt curve analysis of RT-qPCR primers and housekeeping gene validation. **(A)** From the melt curves which plot for the derivative reporter against temperature, we observe that all our RT-qPCR primers were able to produce a single amplification product as seen from the single peaks. **(B)** We verified that two other housekeeping genes (*ACTB* and *HPRT*), when normalized to *GAPDH* mRNA expression, showed no significant difference between the WT and SMA sample groups. SMA: SMA3 and SMA1 iHep lines combined. WT n=9, SMA3 n=9, SMA1 n=9, each with three biological replicates, from three independent experiments. This demonstrates that *GAPDH* mRNA expression between the WT and SMA lines have no significant difference and is appropriate as a housekeeping gene for the context of this study. Data were analyzed using unpaired two tailed Student's t-test. No outliers were detected using the ROUT test with a maximum false discovery rate (FDR) of 1%. Data are presented as mean (\pm SD). ns: non-significant.

Supplemental Figure 3

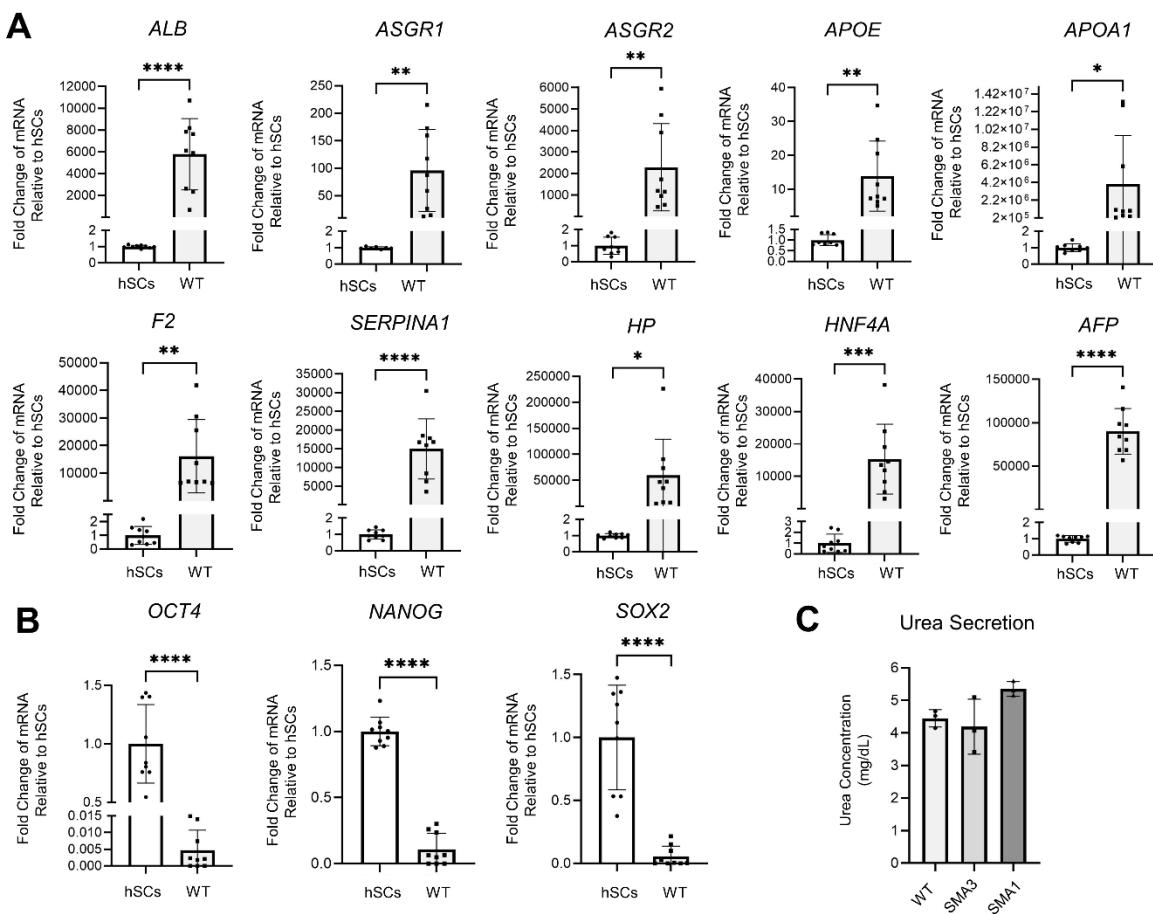


Figure S3. WT iHeps differentiation. (A) RT-qPCR analysis of hepatocyte biomarkers between Day 24 WT iHeps and Day 0 WT stem cells. For *ASGR1*, two outliers from hSCs were removed. For *ASGR1*, WT hSC n=7 and WT iHeps n=9. For all other hepatocyte biomarker genes, WT hSC n=9 and WT iHeps n=9. WT hSCs and WT iHeps data include three biological replicates each, and are from three independent experiments. (B) RT-qPCR analysis of stem cell biomarkers in Day 24 WT iHeps and Day 0 WT stem cells. WT hSCs n=9 and WT iHeps n=9, each with three biological replicates and from three independent experiments.. (A - B) For all RT-qPCR, fold change results were derived using the comparative $\Delta\Delta Ct$ method. (C) Urea expression assay of Day 24 WT, SMA3 and SMA1 iHeps, via urea detection in conditioned cell culture media. WT n=3, SMA3 n=3, SMA1 n=3, (three biological replicates). (A - C) Data sets involving two sample groups were analyzed using unpaired two-tailed Student's t-test, while those with three or more groups were analyzed using One-Way ANOVA test with Tukey's multiple comparison test. Any outliers were detected and removed using the ROUT test with a maximum false discovery rate (FDR) of 1%. Data are presented as mean (\pm SD). * p-value < 0.05; ** p-value < 0.01; *** p-value < 0.001; **** p-value < 0.0001.

Supplemental Figure 4

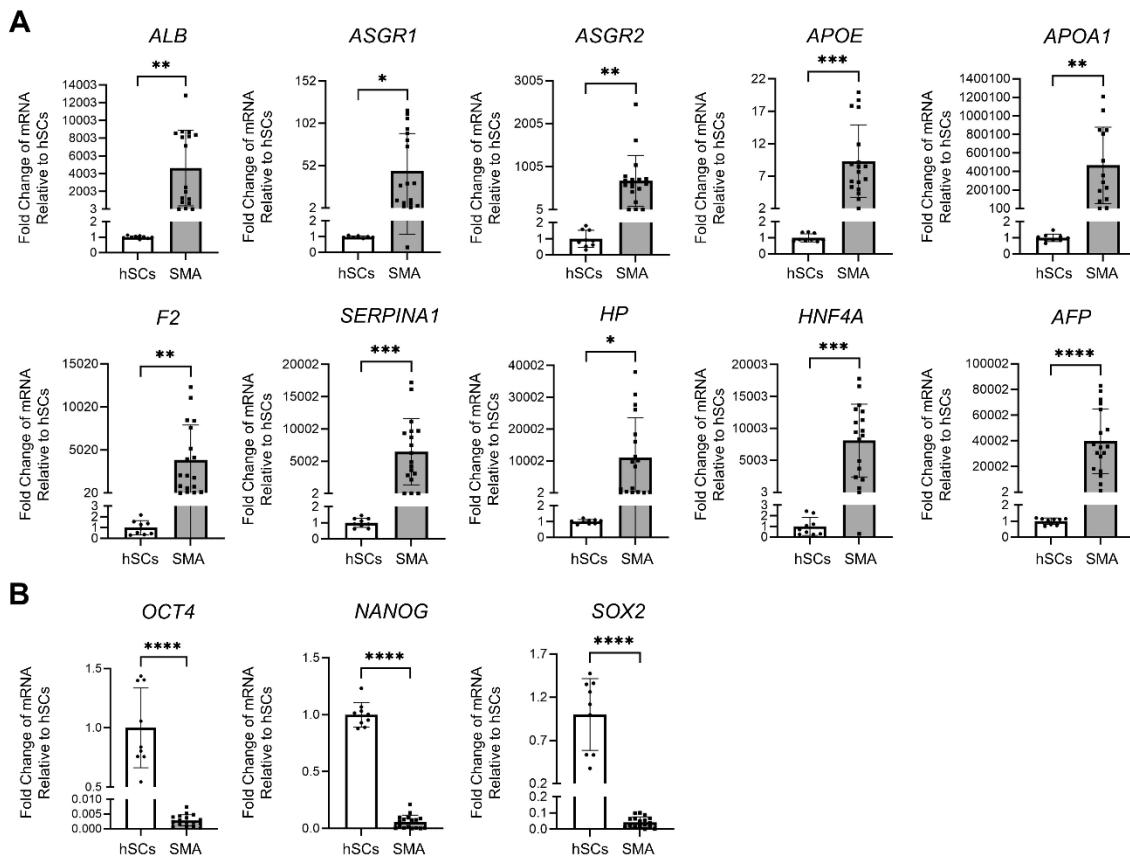


Figure S4. SMA iHeps differentiation. (A) RT-qPCR analysis of hepatocyte biomarkers between the WT stem cells (hSCs) and the SMA iHeps (SMA3 and SMA1 combined). For *ALB*, one outlier from SMA was removed, WT hSCs n=9, SMA3 iHeps n=9, SMA1 iHeps n=8. For *ASGR1*, two outliers from hSCs were removed, WT hSCs n=7, SMA3 iHeps n=9, SMA1 iHeps n=9. For *APOA1*, four outliers from SMA were removed, WT hSCs n=9, SMA3 iHeps n=7, SMA1 iHeps n=7. For all other hepatocyte biomarker genes, WT hSCs n=9, SMA3 iHeps n=9, SMA1 iHeps n=9 (B) RT-qPCR analysis of stem cell biomarkers between the WT stem cells (hSCs) and the SMA iHeps (SMA3 and SMA1 combined). For *OCT4*, *NANOG* and *SOX2*, one outlier from SMA was removed. For *OCT4* and *NANOG*, WT hSCs n=9, SMA3 iHeps n=9, SMA1 iHeps n=8, and for *SOX2*, WT hSCs n=9, SMA3 iHeps n=8, SMA1 iHeps n=9. (A – B) Data is from three independent experiments, each with three biological replicates per sample group. Data was analyzed using unpaired two-tailed Student's t-test. Any outliers were detected and removed using the ROUT test with a maximum false discovery rate (FDR) of 1%. Data are presented as mean (\pm SD). * p-value < 0.05; ** p-value < 0.01; *** p-value < 0.001; **** p-value < 0.0001.

Supplemental Figure 5

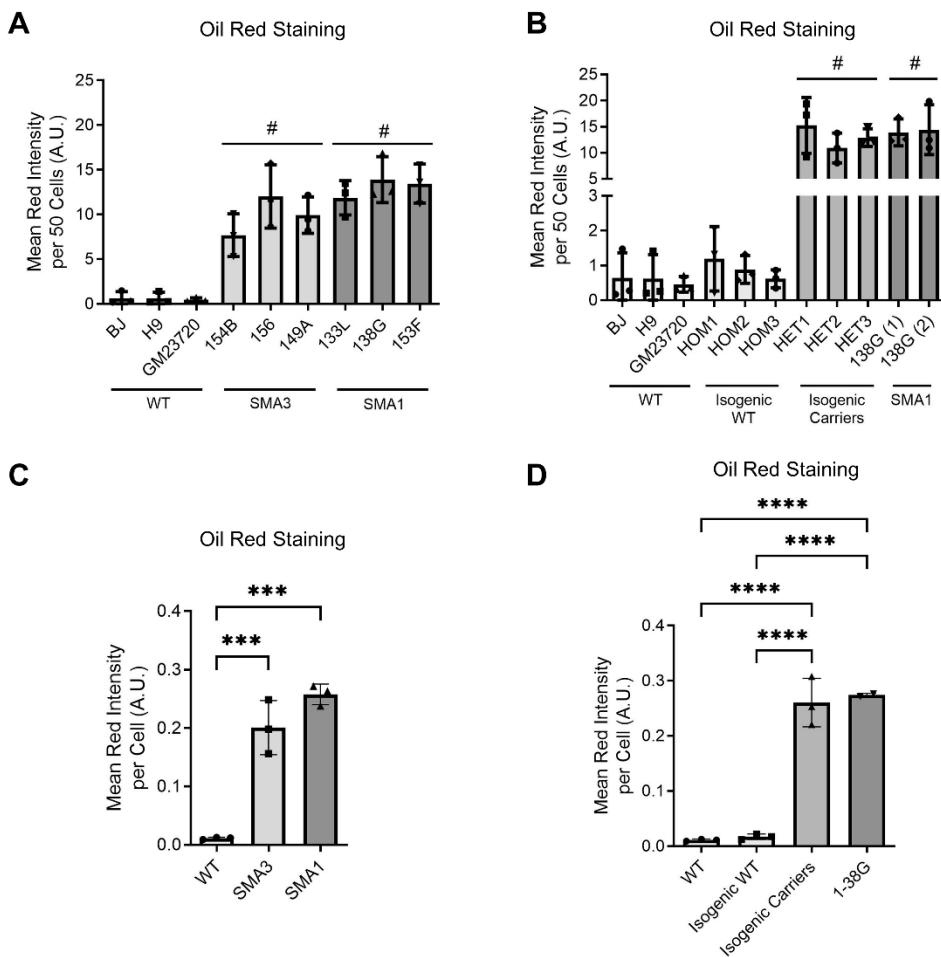


Figure S5. Oil Red O staining of Day 24 WT and SMA iHeps. Oil red staining positively correlates with levels of neutral triglycerides and lipids in iHeps. **(A)** Mean red intensity staining per every 50 cells (A.U.) between the individual WT, SMA3 and SMA1 iHep lines. # represents statistically significant difference between the respective cell lines with respect to WT lines. **(B)** Mean red intensity staining per every 50 cells (A.U.) between the individual WT, isogenic WT, isogenic carriers and 1-38G iHep lines. # represents statistically significant difference between the respective cell lines with respect to WT lines and isogenic WT lines. **(A - B)** For all cell lines, n=3, and are representative of three independent experiments. **(C)** Mean red intensity staining per cell (A.U.) of the respective iHep lines under WT (n=3), SMA3 (n=3) and SMA1 (n=3). Data is representative of three independent experiments. **(D)** Mean red intensity staining per cell (A.U.) of the respective iHep lines under WT (n=3), Isogenic WT (n=3), Isogenic carriers (n=3) and SMA1 (n=2). Data is representative of three independent experiments. **(A - D)** Data were analyzed using One-Way ANOVA test with Tukey's multiple comparison test. No outliers were detected using the ROUT test with a maximum false discovery rate (FDR) of 1%. Data are presented as mean (\pm SD). *** p-value < 0.001; **** p-value < 0.0001.

Supplemental Figure 6

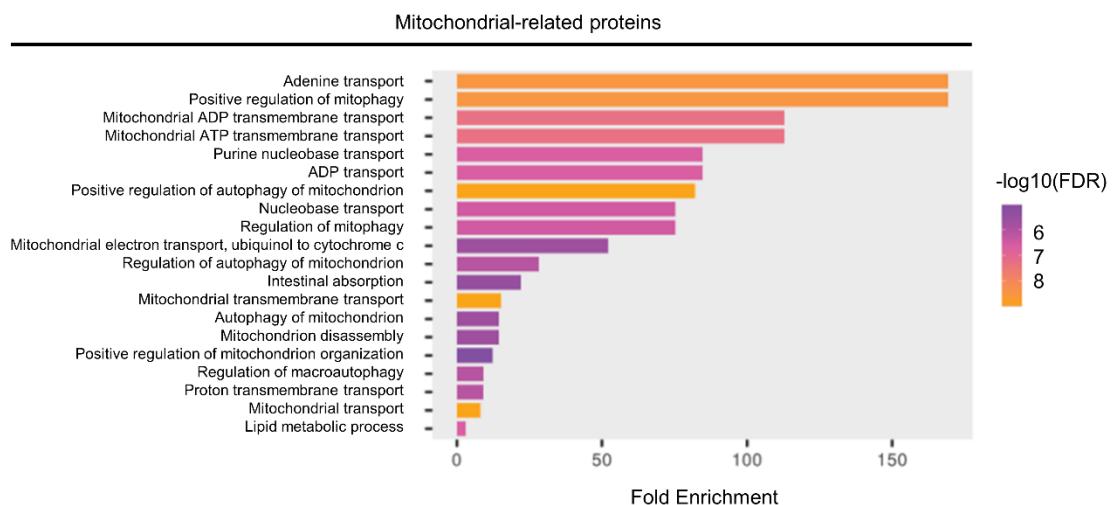


Figure S6. Proteins in Cluster 1, mitochondrial-related processes, subjected to term enrichment analysis using the ShinyGO app. Only the top 20 pathways associated with the cluster are shown and pathways were selected by their $-\log_{10}(\text{FDR})$ scores and sorted by fold enrichment.

Supplemental Figure 7

Golgi apparatus and ER-related proteins

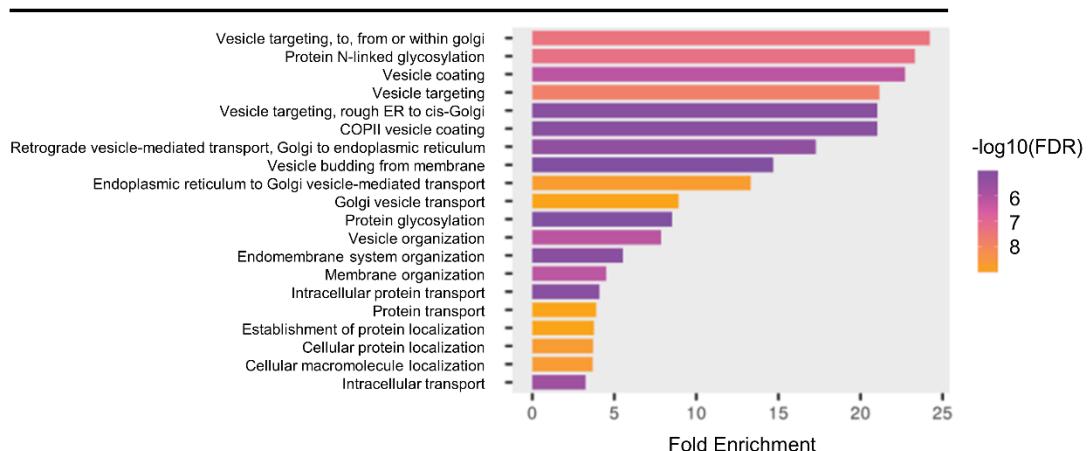


Figure S7. Proteins in Cluster 2, golgi apparatus and endoplasmic reticulum (ER) related processes, subjected to term enrichment analysis using the ShinyGO app. Only the top 20 pathways associated with the cluster are shown and pathways were selected by their $-\log_{10}(\text{FDR})$ scores and sorted by fold enrichment.

Supplemental Figure 8

Protein synthesis and metabolism related

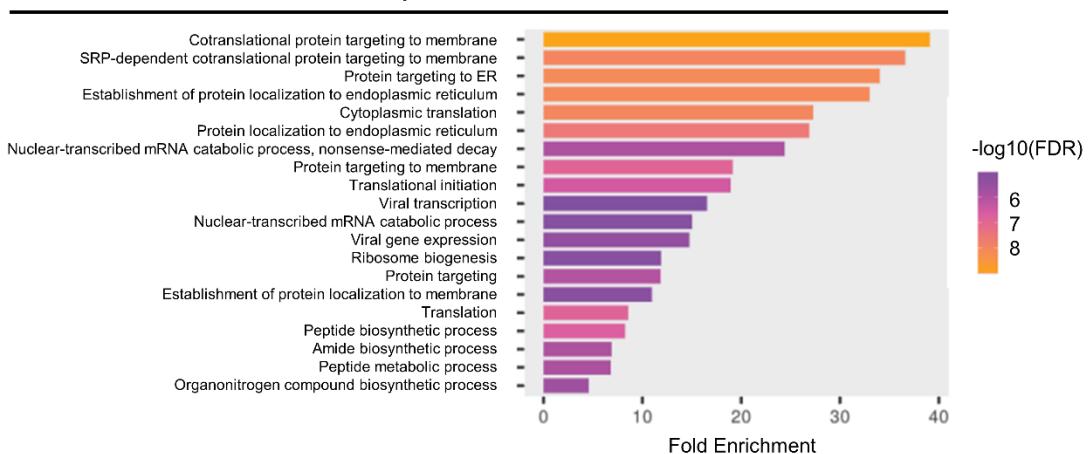


Figure S8. Proteins in Cluster 3, protein synthesis and metabolism related processes, subjected to term enrichment analysis using the ShinyGO app. Only the top 20 pathways associated with the cluster are shown and pathways were selected by their $-\log_{10}(\text{FDR})$ scores and sorted by fold enrichment.

Supplemental Figure 9

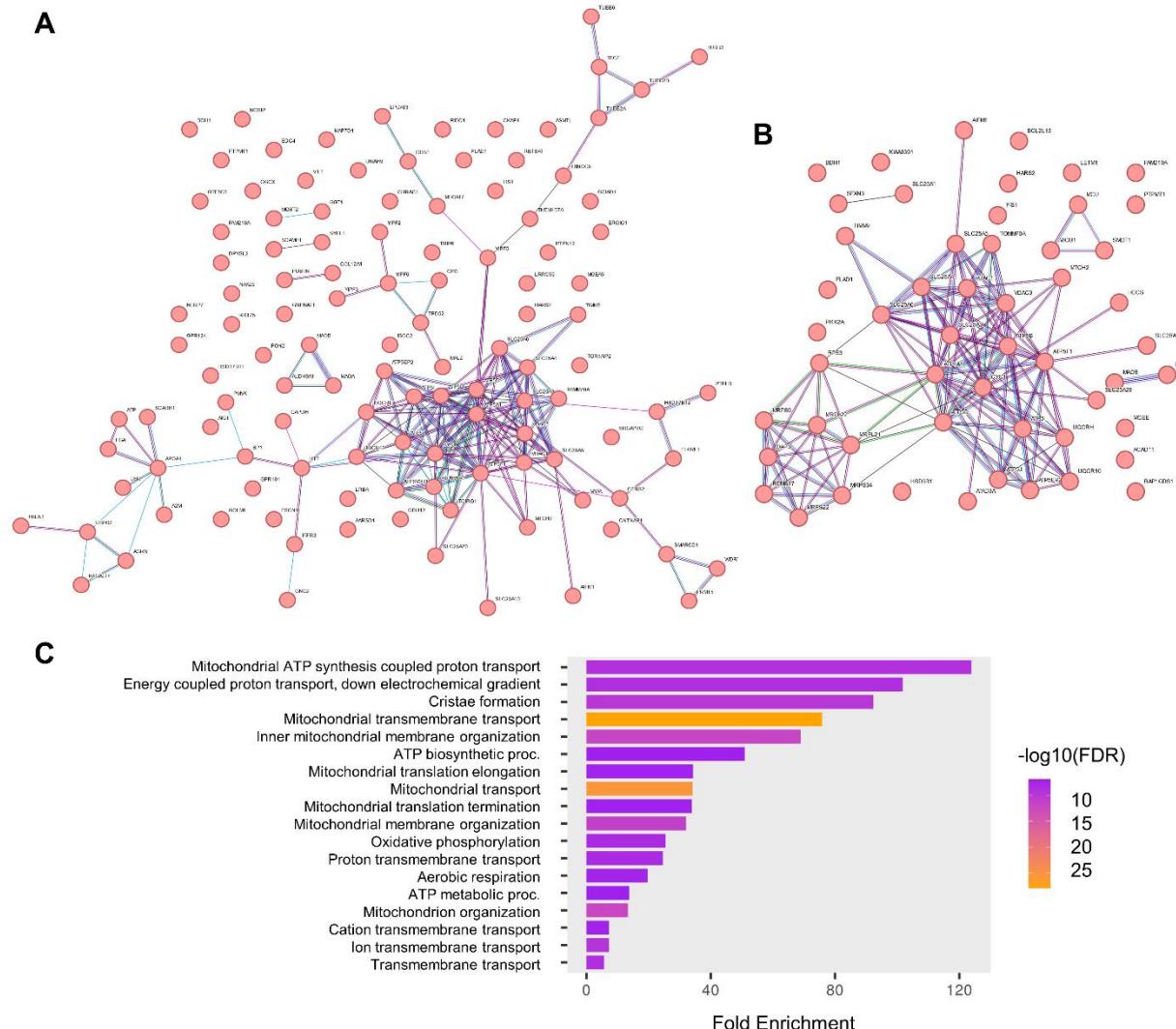


Figure S9. Network analysis of Day 24 WT and SMA iHeps. (A) STRING analysis of proteins in Cluster 1 (red) suggested that there were mitochondrion-related processes in the interaction network. **(B)** STRING analysis of a subset of proteins in Cluster 1 involved in mitochondrial related processes in the interaction network using the UniProt database. **(C)** The top 20 pathways enriched by the subset of proteins in **(B)** selected by their $-\log_{10}(\text{FDR})$ scores and sorted by fold enrichment.

Supplemental Figure 10

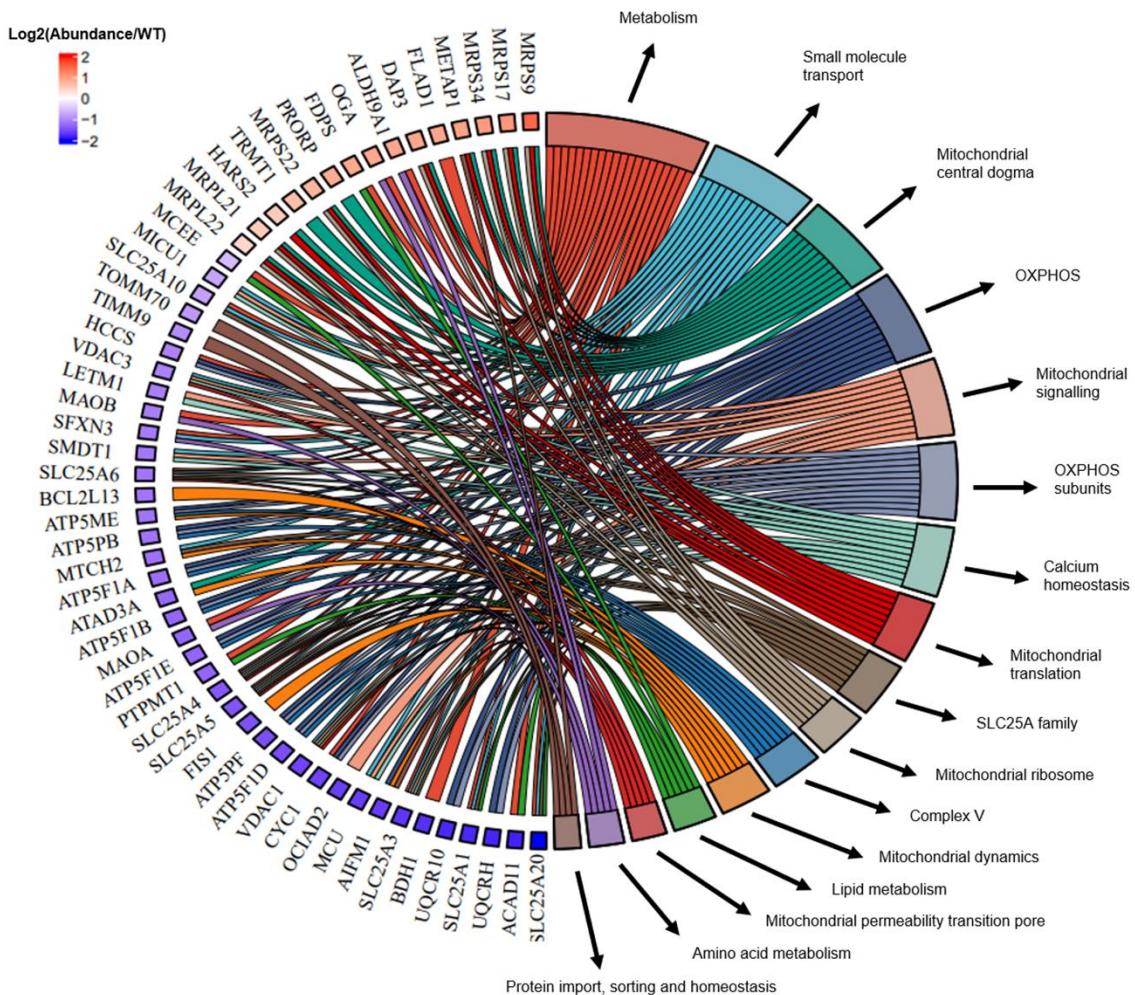


Figure S10. Circos plot associating mitochondrial protein hits with mitochondrial pathways.

Protein hits associated with mitochondria based on the subcellular localization on the UniProt database and mitochondrial pathway using MitoCarta3.0 mitochondrial pathway inventory were used to generate Circos plot.

Supplemental Figure 11

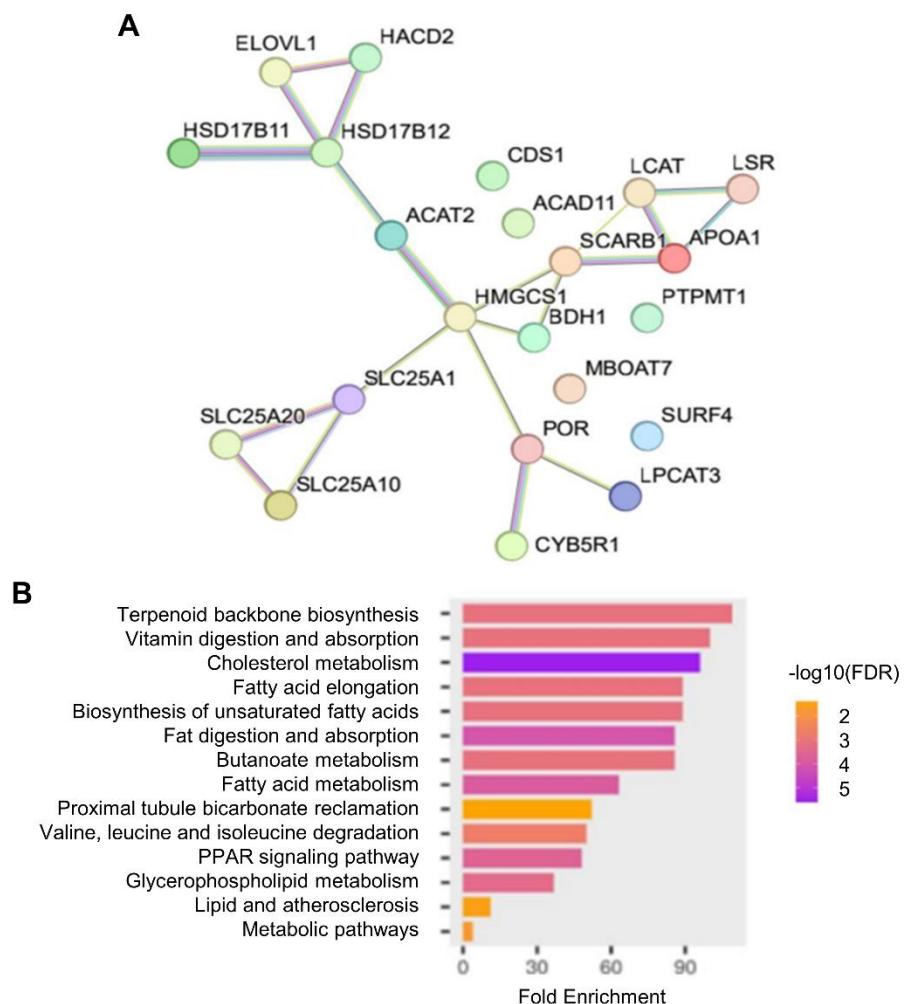


Figure S11. Network analysis of Day 24 WT and SMA iHeps. (A) STRING analysis of proteins involved in lipid homeostasis including lipid transport, triglyceride accumulation, fatty acid oxidation and cholesterol metabolism in the interaction network using the UniProt database. (B) The top 20 pathways were shown out of the top pathways selected.

Supplemental Figure 12

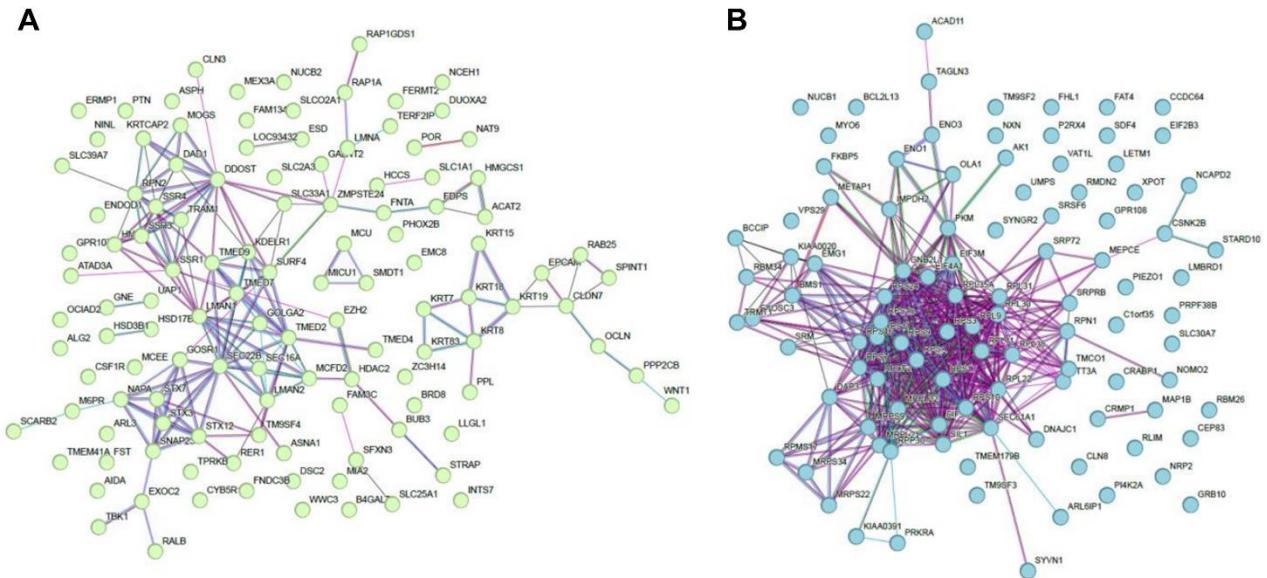


Figure S12. Network analysis of Day 24 WT and SMA iHeps. (A) STRING analysis of proteins in Cluster 2 (green) showed enrichment of golgi apparatus and endoplasmic reticulum related processes in the interaction network using the UniProt database. (B) STRING analysis of proteins in Cluster 3 (blue) showed enrichment of protein synthesis and metabolism related processes in the interaction network using the UniProt database.

Supplemental Figure 13

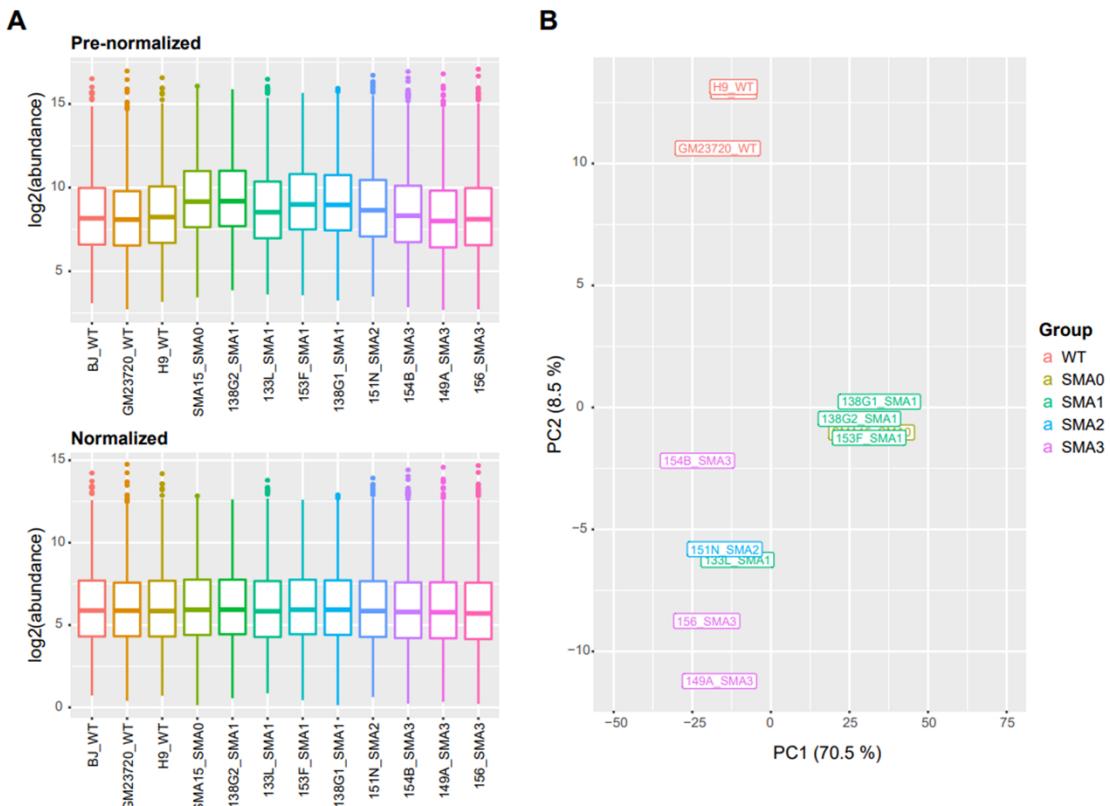


Figure S13. Normalization and principal component analysis (PCA) of Day 24 WT and SMA iHeps. (A) Total intensities for 4,058 proteins identified across all 12 iHeps samples were subjected to median normalization to eliminate potential differences across samples. Results are summarized using boxplots, with each bar representing the interquartile range, the horizontal line within each bar representing the median and the points outside of the whiskers as outliers. Box plots on the top panel show the distribution of the original protein intensities and the bottom panel show the distribution of the protein intensities after the median normalization. (B) Principal component analysis of all iHeps samples post-normalization. The first principal component explains 70.5% of the total data variability, separating the less severe cases (WT, SMA2, SMA3) from the more severe cases (SMA0, SMA1). The second principal component explains 8.5% of the total data variability, separating WT from the SMA samples.

Supplemental Figure 14

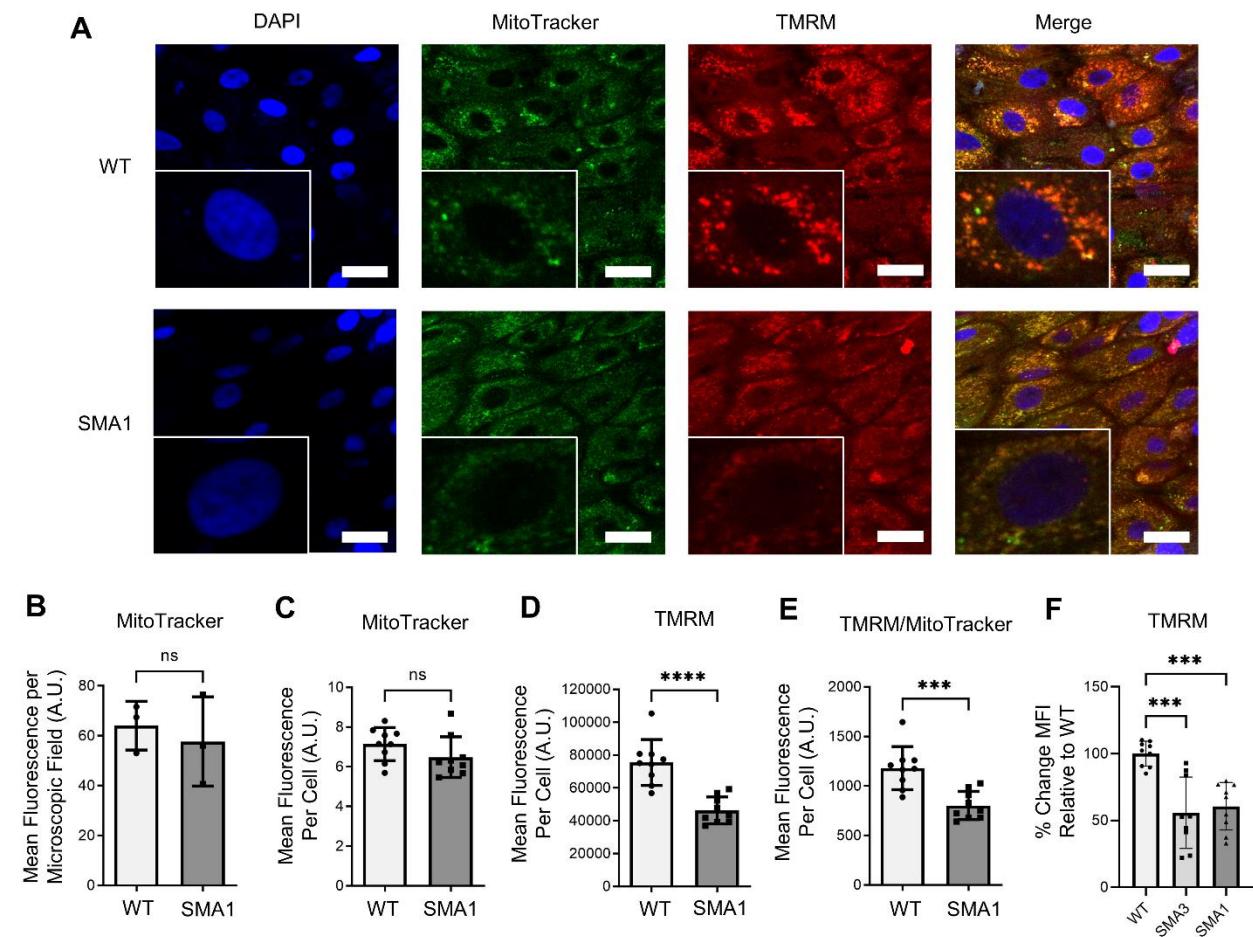


Figure S14. Mitochondrial membrane potential (MMP) reduction in day 24 SMA iHeps is intrinsic to mitochondria. (A) Confocal microscopy analysis of WT (BJ) and SMA1 (GM24468) iHeps. DAPI (blue) stains the nucleus. MitoTracker (green) stains mitochondria. Scale bar = 30μm. TMRM (red) staining positively correlates with MMP levels and co-localizes with MitoTracker staining. WT n=1 and SMA1 n=1. (B - E) ImageJ analysis of confocal staining showing that both TMRM only and TMRM normalized to MitoTracker staining showed reduction in the SMA1 iHep line. Data is from one independent experiment. Data was analyzed using unpaired two-tailed Student's t-test. (B) WT n=3, SMA1 n=3. (C) WT n=9, SMA1 n=9. (D) WT n=9, SMA1 n=9. (E) WT n=9, SMA1 n=9. (F) Mitochondrial membrane potential (MMP) quantification by TMRM, using flow cytometry. WT n=9, SMA1 n=9, SMA3 n=9, each with three biological replicates, and from three independent experiments. Data were analyzed using One-Way ANOVA test with Tukey's multiple comparison test. (B – F) No outliers were detected using the ROUT test with a maximum false discovery rate (FDR) of 1%. Data are presented as mean (\pm SD). *** p-value < 0.001; **** p-value < 0.0001.

Supplemental Figure 15

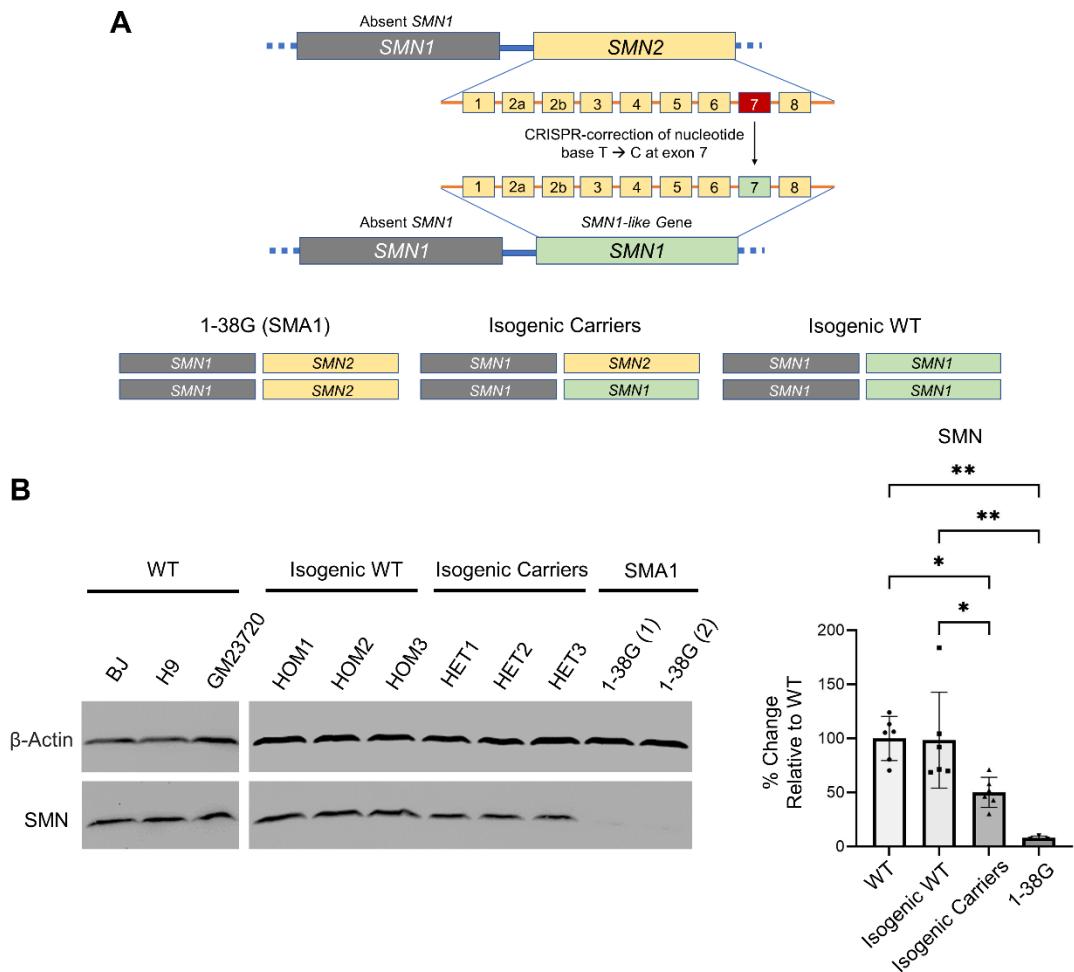


Figure S15. Rescue of SMN expression in Day 24 SMA Type I iHeps with CRISPR-editing.

(A) CRISPR-editing of *SMN2* in the 1-38G (SMA1) cell line into *SMN1*-like genes through a T → C nucleotide change, to investigate SMN protein rescue on an isogenic background. (B) Western blot of SMN protein expression with beta-actin as the housekeeping protein for normalization, showing rescue of SMN protein expression, with the *SMN1*-like gene. Quantification of normalized SMN expression in the iHep lines was done using ImageJ analysis. Data was then presented as percentage levels relative to the mean of WTs. WT n=6, Isogenic WT n=6, Isogenic Carriers n=6, each with three biological replicates, and 1-38G n=3. Data is from two independent experiments. No outliers were detected using the ROUT test with a maximum false discovery rate (FDR) of 1%. Data are presented as mean (\pm SD). * p-value < 0.05; ** p-value < 0.01.

References

1. Stuhr, N. L., Nhan, J. D., Hammerquist, A. M., Van Camp, B., Reoyo, D. and Curran, S. P. (2022). Rapid Lipid Quantification in *Caenorhabditis elegans* by Oil Red O and Nile Red Staining. *Bio-protocol* 12(5): e4340. DOI: 10.21769/BioProtoc.4340.

Q96LZ7	RMDN2	Regulator of microtubule dynamics 2	-0.343678116	0.000347731	-1.26241	-1.21863	-0.33758	-0.60938	3	0.005728 Cluster 3	-0.857000402
Q99571	P2RX4	P2X purinoceptor 4	-0.558813398	0.000580254	-2.19951	-1.65236	0.191818	-0.32408	2	0.006707 Cluster 3	-0.996034054
Q13286	CLN3	Battenin	-0.525373825	0.001705867	-1.97697	-1.78347	-0.67293	-0.81963	1	0.009859 Cluster 2	-1.3132495
P07333	CSF1R	Macrophage colony-stimulating factor 1 receptor	-0.726239013	0.000114023	-2.63725	-2.02755	-0.77689	-0.04753	2	0.004443 Cluster 2	-1.372303254
P0DJ0	SRGAP2C	SLT-ROBO Rho GTPase-activating protein 2C	0.38465523	0.000944591	1.407452	1.036007	0.314476	-0.05918	2	0.007715 Cluster 1	0.674688565
P19012	KRT15	Keratin, type I cytoskeletal 15	-0.615390351	0.000297057	-2.32786	-1.98101	-0.96121	-0.75138	1	0.005576 Cluster 2	-1.505364127
Q9NVW2	RLIM	E3 ubiquitin-protein ligase RLIM	-0.692148667	0.000812938	-2.36608	-1.98175	-0.33274	0.037637	1	0.007396 Cluster 3	-1.160734104
Q9Y5Q8	GTF3C5	General transcription factor 3C polypeptide 5	0.483137045	0.00072881	1.6509	1.247392	0.229822	-0.32927	2	0.00712 Cluster 1	0.699711258
Q709F0	ACAD11	Acyl-CoA dehydrogenase family member 11	-0.472447568	0.000759601	-1.88073	-1.60564	-0.51418	-0.83397	2	0.007162 Cluster 3	-1.208630586
P08727	KRT19	Keratin, type I cytoskeletal 19	-0.717632167	0.0002446	-2.56283	-2.22311	-0.75816	-0.49073	1	0.005213 Cluster 2	-1.508705883
Q9UHD2	TBK1	Serine/threonine-protein kinase TBK1	-0.591776096	0.001623385	-2.50723	-1.90957	-0.91576	-1.00045	1	0.009613 Cluster 2	-1.583251059
Q13509	TUBB3	Tubulin beta-3 chain	0.343737359	0.000539517	1.699836	0.970437	0.477184	0.510202	2	0.006627 Cluster 1	0.914414615
P08729	KRT7	Keratin, type II cytoskeletal 7	-0.689535864	0.000871646	-2.59922	-2.3964	-0.78064	-1.19255	1	0.007557 Cluster 2	-1.742201966
Q92508	PIEZ01	Piezo-type mechanosensitive ion channel component 1	-0.664308811	0.000424136	-2.03076	-1.95777	-1.36885	0.035834	2	0.006001 Cluster 3	-1.330383601
Q9BVA1	TUBB2B	Tubulin beta-2B chain	0.624279425	0.001016797	2.314391	1.634082	0.012088	-0.2264	2	0.00785 Cluster 1	0.933539355
Q9Y592	CEP83	Centrosomal protein of 83 kDa	-0.992400578	0.00042139	-4.0305	-2.93619	-0.2449	-0.76885	2	0.005987 Cluster 3	-1.995109904
Q96P66	GPR101	Probable G-protein coupled receptor 101	0.494300065	0.001084452	1.724625	1.421086	0.765019	0.079517	2	0.007998 Cluster 1	0.997561579
Q9COA0	CNTNAP4	Contactin-associated protein-like 4	0.473106242	0.000143203	1.885574	1.535457	0.684155	0.695224	2	0.004614 Cluster 1	1.200102403