## EXTENDED MATERIALS AND METHODS

### *Cell Culture Techniques and CRISPR Homology Directed Repair*

H1 hPSC (WA01, WiCell) were used in this publication to generate the rtTA Control and TET-On hPSC lines. HPSC's were seeded on Matrigel (Corning), cultured using mTESR Plus (Stem Cell Tech), passaged using ReleasR (Stem Cell Tech), and frozen down in mFRESR (Stem Cell Tech). During passaging and thawing, hPSCs were cultured in Rock Inhibitor (Stem Cell Tech).

Generation of a CRISPR genetically engineered TETon MEGF10 overexpression cell line began with the first insertion of the EF1A-rtTA-mTagBFP2 transgene into the AAVS1 genomic safe harbor. We used a single guide RNA (IDT) bound to a purified CAS9 protein (IDT) to cleave the *AAVS1* site of insertion, using nucleofection as the method of plasmid and gRNA delivery (Lonza). We then used PCR screening to verify the presence of the correct insertion in the parental bulk cell sample before harnessing the mTagBFP2 fluorescence to single cell plate BFP+ cells for culturing and subsequent clonal line PCR screening. We obtained several rtTA clonal insertion lines, and the line we moved forward with for the secondary transgene insertion successfully passed karyotyping assays for genetic abnormalities. Sanger sequencing of PCR products verified that the correct transgene insertion was incorporated at the AAVS1 locus. This line was used as the rtTA Control line through this study.

We then repeated the process to generate a secondary transgene insertion, incorporating the TET-MEGF10-TdTomato into the *ROSA26* locus. We again used PCR screening to verify that the transgene was inserted to the right position and to verify that the insertion was homozygous. With another round of single cell plating and screening through surviving colonies of a 96-well plate, we then split the colonies and treated one replicate with 2μM doxycycline. We moved forward with the colonies that responded to treatment with a clear fluorescent RFP+ signal. It should be noted that after clonal single-cell plating and expansion, ~30-50% of Tet-On hPSCs responded to treatment – we hypothesize that this may be due to genetic silencing of the

transgenes introduced. Thus, we gently pulsed with 1μM doxycycline for ~24 hours and FACS enriched for the tdTomato+ hPSCs. These enriched hPSCs were ~100% responsive when treated with doxycycline two passages later, and thus we proceeded with directed differentiation with these hPSCs.

Patient-modeling EMARDD mutations were also generated using CRISPR genome engineering. We used the rtTA Control line described above as a template for EMARDD disease modeling lines to minimize genetic line-to-line variability. sgRNA, CAS9 protein (IDT), and template plasmid were nucleofected into hPSCs and single cell clonal insertions were screened from this parental reaction. The EMARDD-Logan line was replicated with an identical, 10-bp duplication event as described in Logan 2011's work diagnosing a novel EMARDD patient variant. This genetic mutation gives rise to an early stop codon, but importantly, it likely still produces a truncated protein since the mutation occurs past halfway in the coding sequence of MEGF10. The EMARDD-Takayama line was not an identical mutation event; rather than the 2-bp deletion, we introduced a 1-bp insertion that introduced an identical frameshift mutation, resulting in a premature stop codon, with only 10 base pairs that differ between our model and the patient genotype. Both lines were applied in our studies, however in engraftment studies, only the EMARDD-Logan line was included.

For directed differentiations, we followed our Lab's publicly available protocol of skeletal muscle derivation from hPSCs. For FACS enrichments of SMPCs after directed differentiation, we used HNK1-FITC (BioLegend), ERBB3-PECy7 (BioLegend), NGFR-APC (BioLegend), and AlexaFlour780 Live Dead Viability stain (Fisher) to accommodate endogenous BFP and tdTomato. We opted to pulse Tet-On differentiations with doxycycline ~24 hours prior to sorting to enrich a tdT+ population that is highly responsive to treatment for engraftment studies. SMPCs were plated at ~150-300K per well of a 6-well plate and cultured in SKGM2 media (Lonza) + bFGF. Generally, they took ~3-4 days to recover and reach 75% confluency before passaging/freezing. *In vitro* myogenic differentiation and fusion assays were generated using N2

differentiation media; 1.2% N2 supplement (Fisher), 1.2% Immuno transferrin selenium (Fisher), IGF1 (Stem Cell Tech), in 1:1 DMEM-F12 media. Skeletal muscle progenitor cells were grown in SKGM2 media (Lonza), passaged in Tryple (Fisher), and frozen in CryoStor (Stem Cell Tech).

***RNA Sequencing of SMPCs***

We established directed differentiation cultures of rtTA Control and Tet-On hPSC lines. At ~day 40 of differentiation, the TETon cultures were treated with 2μM doxycycline (in DMSO; Stem Cell Tech) to pulse MEGF10 overexpression, and the rtTA Control line was treated with equal amounts of DMSO (1:1000 in N2 media). At day 47, rtTA Control SMPCs were collected (n=5) by enriching live HNK1-, ERBB3+, NGFR+ and TETon SMPCs were collected (n=4) via live HNK1-, ERBB3+, NGFR+, tdTomato+ SMPCs. Different wells of a 6-well plate were handled separately to serve as biological replicates. FACS enriched SMPCs were pelleted, and RNA was immediately harvested using the RNA-Easy MicroPrep Kit with gDNA elimination (Qiagen). These RNA samples were used to construct libraries for Illumina sequencing with the assistance of the UCI GHTF Core (SMARTseq library construction kit). We sequenced ~60million reads and FastQ files were groomed user Fastq-groomer and aligned using HiSAT2 (*Homo Sapiens* hg38 GTF reference file) and processed in *Galaxy* to generate raw count matrices. Samples were. These matrices were used as input for Deseq2 analysis in R Studio for data visualization and downstream analysis. We closely followed DESeq2's platform as guided by publications and github resources to generate PCA clustering, Volcano plot, and Violin plot generation across our normalized gene expression data. DESeq2 uses a Fisher Exact Test combined with a Multiple Testing Correction (Benjamini-Hochberg) to generate adjusted p values that correct for over prediction of significantly upregulated DEGs.

***Skeletal Muscle Stem Cell Engraftments***

~6-10 week old (MDXD2/WT) Pax7-Cre ROSA-DTA inducible ablation mice were treated with tamoxifen chow (Envigo) for at least 7 days prior to localized muscle injury and engraftment. For the Tet-On engraftments, doxycycline chow (Envigo) was administered. To label cell

proliferation *in vivo*, EDU (Sigma) was administered to mice in water at 0.5mg/mL. Live HNK1-, ERBB3+, NGFR+ SMPCs were enriched after directed differentiation and plated on Matrigel to recover. After ~3 days, 75% confluence SMPCs were passaged ~1:5 to encourage expansion of the cells to maximize engraftment inputs. After another ~3 days, SMPCs were pelleted and resuspended in HBBS without Ca/Mg (Fisher) at a concentration of 5μL/million SMPCs for engraftment. ~1.2 million SMPCs were engrafted into the Tibialis Anterior of mice, and importantly, these muscles had been subjected to 50μL of 1.2% Barium Chloride (Sigma) injury ~24hours prior.

After 30 days, mdx-NSG mice were euthanized and TA muscles dissected. The TA muscles were immediately embedded in optimum cutting temperature (OCT) compound and flash-frozen in isopentane cooled by liquid Nitrogen. Embedded muscles were stored at -80°C until sectioned in 9-10μm slices using a Leica microcryotome (LSXII). 1200μm of TA muscle were removed and 20 sections collected on positively charged frosted microscope slides in 50μm intervals, and 300μm of tissue were collected for RNA or western blots. In this manner we calculated that each quadrant contained 1300μm of muscle tissue. Quadrants were collected until the entire TA was sectioned. Thus, we could determine the position of each cross section for downstream analysis of cell engraftments. Microscope slides were kept in aluminum foil covered boxes at -20°C.

***Real Time PCR***

Primers were designed using NCBI primer blast or based on published work. Primers (10μM) were validated using cells known to express the following genes. CDNA concentrations were 5-fold serially diluted, 5ng/μL to 0.008ng/μl. Primers with 0.9-1.1 efficiency were used for experiments.

***Immunofluorescent staining***

To perform staining, sections on slides were thawed from -20°C freezers (45 minutes to promote PAX7 antigen binding) and washed in PBS containing 0.2% tween-20 (PBST, 7.4pH).

Slides were fixed in 2% PFA for 10 minutes and washed with PBST. Sections were then blocked in 0.2% gelatin, 3% BSA and 0.1% Tween, 10% goat serum for 90 minutes at room temperature. Sections were immunostained with human-specific antibodies to h-Spectrin (1:80; Vector laboratories), h-Lamin AC (1:120; Fisher) and h-Dystrophin (1:6; MANDYS106) to positively identify engrafted human cells. Human/mouse-specific laminin (1:80; Vector laboratories) was used to denote the basal lamina of all myofibers. Sections were alternatively stained with h-PAX7 (10μg/ml; DSHB), h-Lamin AC, h-Spectrin and Laminin to identify human SMPC/SCs residing in the satellite cell position.

SMPC/SCs and myotubes were washed in PBS and fixed in 4% PFA for 15 minutes. Cells were permeabilized with 0.3% Triton-X for 5 minutes. Cells were then blocked in 10% goat serum, 3% BSA for 1 hour. Primary antibodies were added to 2% goat serum in PBS and incubated on cells overnight at 4C. Primaries were washed three times using 0.1% tween-20 in PBS and secondary antibodies in PBS were added for 65 minutes. Secondary antibodies were washed, and cells were counterstained with 1:1000 Hoechst. Cells were imaged using a Zeiss inverted fluorescent microscope.

***Engraftment Quantifications and Statistical Analysis***

Engraftment quantifications of PAX7+ SMPCs and myofibers were collected per cross sectional area. This was conducted by determining the ideal engraftment quadrants and manually counting human PAX7+ nuclei and human Spectrin+ myofibers across the top 2-3 quadrants with the most engraftment. As these data are already normalized according to the cell amounts engrafted (1.2million per engraftment), we did not normalize these quantifications. Statistics were conducted using one-way ANOVA across comparisons with more than 2 samples, and a student's T-test was applied to comparisons between 2 groups.

For Figure 1, we transplanted n=3/4 mice at each timepoint day 1, 5, 10, 20. For Figure 2/3, we transplanted n=3 of each time point across both healthy and dystrophic mice. In Figure 6, we transplanted n=3 for each group of control, EMARDD, and MEGF10 overexpression. Lastly,

for Figure 7, we transplanted n=3/4 mice for control, EMARDD, and Tet-On. In all cases that required spatial transcriptomics, i.e. Figures 1, 2, 3, and 7, we moved forward with spatial transcriptomics on one sample from each time point for cost effective analysis.

***GeoMX DSP Spatial Transcriptomic Analysis***

Workflow:

We proceeded with spatial transcriptomic analysis using the Nanostring/Bruker GeoMX Digital Spatial profiler on engraftment samples relevant for Figures 1,2,3, and 7. The kit that was used was the Whole Transcriptome Human Probe Atlas. We opted for the samples that had the best engraftment and proceeded to conduct spatial transcriptomic analysis on them immediately. While flash freezing tissues does not decrease RNA quality, we anticipated degradation after further storage of sectioned slides at -20C. Thus, we stored freshly sectioned slides for ~1-2 days at 4C and began the spatial workflow as soon as possible. We proceeded with one tissue slide consisting of 3-4 tissue slices per sample, for biological replicates of N=1, but technical replicates of N>10 per myogenic population.

This workflow closely followed the detailed protocol in Clock, Hicks *Methods in Molecular Biology* 2025. To conduct this analysis, engrafted tissue samples were collected by harvesting the engrafted mouse TA and flash freezing in liquid nitrogen. These frozen samples are stable at -80C, with RNA integrity and quality preserved. Frozen engrafted samples were sectioned transversely along the entirety of the tissue to determine where the best region of engraftment presided. We conducted an overnight immunostain to determine where the engraftment was most successful using human lamin AC (Fisher), human spectrin (Fisher), PAX7 (DSHB), and Laminin (Thermo). Critically, during this overnight stain, excess slides were temporarily stored at 4C in a dry box to avoid water build up to preserve RNA integrity and avoid any freeze/thaw degradation of RNA. After determining which region of engraftment was most successful, we moved forward with those specific slides from 4C for spatial transcriptomic analysis. Samples were fixed with 4% formaldehyde then treated with 0.025µg/mL proteinase K to expose RNA binding epitopes. Then,

engrafted tissues were incubated with Human Whole Transcriptome Atlas Probes (Bruker) overnight at 37C. The following day the probes were washed off with formalin (Sigma) and 2xSSC Buffer prior to staining the tissues with morphology markers / antibodies of interest – human lamin AC, human spectrin, PAX7, MEGF10, Laminin, MYOG. The GeoMX DSP Platform was then implemented to soundly image the stained samples for precise and controlled region of interest (ROI) selection. We captured between 5-15 nuclei in each ROI to ensure enough transcriptomic content for downstream analysis.

For Figure 1's engraftment trajectory analysis and tracking of myogenic population changes across time, we were unable to capture all myogenic populations desired. We noted a difficulty in capturing enough engrafted cells, that we were unable to stratify populations into the depth and detail drafted in our hypotheses. For example, PAX7 cells are increasingly rare across time (**Figure 1B**), and while we could capture separate PAX7+MYOD1- and PAX7+MYOD1+ populations at Day 1, we were unable to do so at all time points due to the low number of PAX7 cells supported after engraftment. Thus, we systematically merged our dataset to focus on cells that were collected across several time points reliably.

Sequencing:

After collection of ROIs on the GeoMX DSP, libraries were prepared according to manufacturer workflows, harnessing a seq-code plate that barcodes ROIs from the same collection plate to allow for sample pooling when sequencing. Illumina sequencing at 2 million reads per micron$^2$ (we doubled what the workflow recommended for additional depth) yielded FastQ files, FastQ indices, and a configuration.ini file. The FastQ and .ini file were converted to DCC files using the Linux command line interface, *geomxngspipeline*. These DCC files were reuploaded to the GeoMX DSP interface platform for analysis.

Sample Flagging and Filtration:

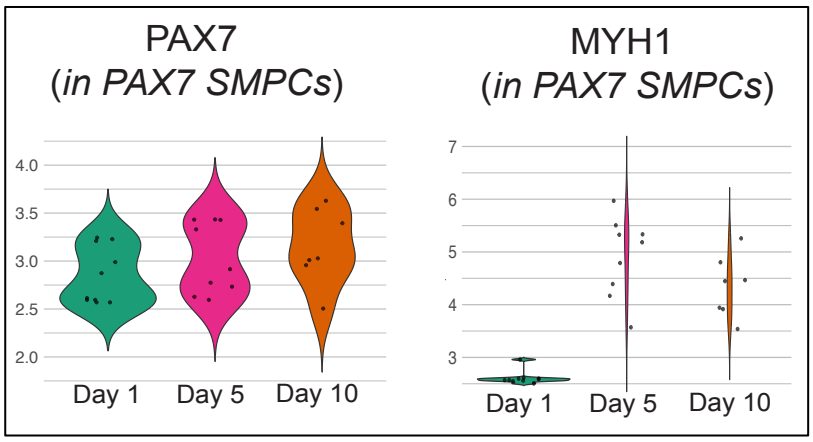
With the uploading of DCC files for sample analysis, we began by downloading unfiltered, unflagged counts matrices for quality control purposes. We assessed filtration cutoffs flagging

samples that had >90% of genes that were expressed below negative probe levels and removed them. For PAX7 samples, we flagged samples that had a ratio of PAX7 transcript / negative probe transcript below 0.75. These flagged samples were removed from analysis prior to filtration of the gene expression profiles. Filtration was conducted at 2% LOQ levels on the GeoMX platform, and we subsequently corrected this filtered dataset to remove genes that were, on average, below the average negative probe count. This dataset was used for analysis and uploaded into R for DESeq2 rigor.

Analysis:

Deseq2 required an input of un-normalized data – thus, we relied on the rigor of Deseq2's build in normalization method which is the Wald Test of Negative Binomial Generalized Linear Model. The Benjamini-Hochberg model is subsequently applied to generate adjusted p-values, and in this study, we considered padj < 0.1 as significant, which is in line with what other publications conducting DESeq2 have applied. These normalized data were then transformed to stabilize variance across samples and datasets using VST, and these values were used to conduct principal component analysis, and project heatmap and violin plots.

In Figure 1, we investigated the global response to the *in vivo* microenvironment by identifying shared hits across time in SMPCs and myocytes, but we first noticed infiltrating myofiber genes in PAX7 SMPCs, as marked by an increase in *MYH1* (below). The biological positioning of myofibers directly adjacent to SMPCs and myocytes led to this technical challenge of spatial transcriptomics. The biological positioning of myofibers directly adjacent to SMPCs and myocytes led to this technical challenge of spatial transcriptomics. We resolved this challenge by identifying myofiber infiltrating genes within our SMPC populations through systematic comparison of the two populations across time and flagging all myofiber differentially expressed genes (DEGs) when analyzing SMPCs across time.




**Infiltration of myofiber transcripts in PAX7 cells**: Violin plot of normalized gene expression values after variance stabilization transformation across time in PAX7 SMPC populations; PAX7 transcripts (left) and MYH1 transcripts (right). Statistical significance determined through DESeq2's metric of Negative Wald Binominal and Multiple Testing Corrections (Benjamini Hochberg) – no significance determined in PAX7, but in MYH1 transcripts Day 1 is significantly lower than other Day 5 and Day 10.