

1 **Primary human myoblasts display only minor alternative polyadenylation**
2 **compared to the transformed C₂C₁₂ model of muscle differentiation.**

3 **Running head:** Alternative polyadenylation is minor in human myotubes.

4 Akriti Varshney¹, Paul F. Harrison^{1,4}, Angavai Swaminathan¹, Sarah E. Alexander²,
5 Bernhard Dichtl³, Séverine Lamon², Traude H. Beilharz^{1*}

6

7 ¹Development and Stem Cells Program, Monash Biomedicine Discovery Institute
8 and Department of Biochemistry and Molecular Biology, Monash University,
9 Melbourne, VIC 3800, Australia.

10 ²Institute for Physical Activity and Nutrition, School of Exercise and Nutrition
11 Sciences, Deakin University, Geelong, Victoria 3220, Australia.

12 ³School of Life and Environmental Sciences, Deakin University, Geelong, Victoria
13 3220, Australia.

14 ⁴Monash Bioinformatics Platform, Monash University, Melbourne, VIC 3800,
15 Australia.

16

17

18

19 Traude Beilharz
20 Development and Stem Cells Program,
21 Monash Biomedicine Discovery Institute
22 and Department of Biochemistry and Molecular Biology,
23 Monash University, Melbourne, VIC 3800, Australia
24 Email: traude.beilharz@monash.edu
25 Phone: +61 3 9902 9183

Varshney et al.

26 **Abstract**

27 Alternative polyadenylation has been linked to multiple developmental and disease
28 transitions. The prevailing hypothesis being that differentiated cells use longer 3'
29 UTRs with expended regulatory capacity whereas undifferentiated cells use shorter
30 3' UTRs. Here, we describe the gene expression and alternative polyadenylation
31 profiles of human primary myoblasts over a time course of differentiation. Contrary to
32 expectations, only minor changes in the 3' end choice were observed. To reconcile
33 this finding with published research, we devised a new bioinformatic method to
34 compare the degree of alternative polyadenylation in the differentiation of primary
35 human and immortalized murine (C₂C₁₂) myoblasts. Differentiated human primary
36 myotubes display only half the alternative polyadenylation of the mouse model, with
37 less than 1/10 of the genes undergoing alternative polyadenylation in C₂C₁₂ cells
38 showing evidence of alternative processing in human primary muscle differentiation.
39 A global reduction in the expression of cleavage and polyadenylation factors in
40 C₂C₁₂, but not in primary human myotubes may explain the lack of alternative
41 polyadenylation in this system. Looking more broadly at transcriptome changes
42 across differentiation shows that less than half of the genes differentially expressed
43 in the immortalized model were recapitulated in primary cells. Of these, important
44 metabolic pathways, such as glycolysis and sterol biosynthesis, showed divergent
45 regulation. Collectively, our data caution against using immortalized cell lines, which
46 may not fully recapitulate human muscle development, and suggest that alternative
47 polyadenylation in the differentiation of primary cells might be less pronounced than
48 previously thought.

49 **Keywords:** alternative polyadenylation; cholesterol biosynthesis; primary human
50 myoblast differentiation

Varshney et al.

51 **Introduction**

52 Alternative polyadenylation (APA) is a co-transcriptional process that generates
53 mRNA isoforms with different 3' untranslated regions (UTRs) from the same genetic
54 locus. The cleavage and polyadenylation (CPA) machinery responsible for APA
55 consists of several evolutionarily conserved protein complexes, including cleavage
56 and polyadenylation specificity factor (CPSF), cleavage stimulation factor (CstF),
57 Cleavage factor I and II (CFI and CFII, and a poly(A) polymerase (Giammartino et al.
58 2011; Tian and Manley 2017). The CPSF component is primarily responsible for
59 detecting a canonical AAUAAA poly(A) signal (PAS) in the 3' UTR, or related non-
60 canonical PAS for downstream endonucleolytic pre-mRNA cleavage and the non-
61 templated addition of adenosine nucleotides for poly(A) tail synthesis (Giammartino
62 et al. 2011; Tian and Manley 2017). When more than one poly(A) signal exists, the
63 choice is regulated by multiple intersecting mechanisms, including the strength of
64 PAS and the relative concentration of the cleavage and polyadenylation machinery
65 (Gruber et al. 2014). The length of the 3' UTR is significant because it can contain
66 regulatory elements for nuclear export, stability, and spatiotemporal translation of
67 mature mRNA (Turner et al. 2018; Beilharz et al. 2019; Pereira-Castro and Moreira
68 2021). Seminal early work showed that 3' UTR shortening resulted in aberrant
69 expression of key oncogenic genes by loss of crucial RNA-binding protein and
70 microRNA binding sequences (Sandberg et al. 2008; Mayr and Bartel 2009). The
71 use of proximal PAS, and thus shorter 3' UTRs, has now been associated with
72 proliferative growth in multiple stem cell and cancer models (Kandhari et al. 2021;
73 Yuan et al. 2021).
74 The differential expression of CPA factors can explain the relationship between
75 proliferation and APA (Gruber et al. 2014). For example, altered expression of

Varshney et al.

76 CFIm25 and PCF11 has been linked to 3' UTR shortening and tumorigenesis in
77 glioblastoma (Masamha et al. 2014) and neuroblastoma (Ogorodnikov et al. 2018).
78 In contrast, distal poly(A) site selection and longer 3' UTRs have been widely
79 reported in cell differentiation (Ji et al. 2009; Grassi et al. 2019; Sommerkamp et al.
80 2021). For example, high levels of the polyadenylation factor Fip1 promote the
81 shortening of 3' UTRs essential for embryonic stem cell maintenance and
82 proliferation (Lackford et al. 2014). Similarly, the differentiation of immortalised
83 skeletal myoblasts leads to the loss of PCF11 with global 3' UTR lengthening (Wang
84 et al. 2019). APA has also been reported in multiple models of cardiac myotube
85 differentiation and disease (Ji and Tian 2009; Nimura et al. 2016; Soetanto et al.
86 2016; Yang et al. 2022).

87 Skeletal muscle development, or myogenesis, is a complex and highly regulated
88 cellular process and has been the model of choice for investigating the role and
89 regulation of APA in cell differentiation (Li et al. 2015; Wang et al. 2019). The model
90 relies on a population of quiescent muscle stem cells, termed myoblasts or satellite
91 cells, that exit their proliferating state and undergo terminal differentiation
92 (Benzinger et al. 2012). Differentiating muscle cells or myotubes then fuse to
93 become multinucleated myofibers (Chal and Pourquié 2017; Jiwlawat et al. 2018),
94 the contractile units that allow for muscle movement in adults (Korthuis 2011). Some
95 myoblasts retain their stem properties in adult tissues, allowing postnatal skeletal
96 muscle growth and regeneration (Charge and Rudnicki 2004). One of the most
97 common models of mammalian muscle development used for transcriptomic and
98 metabolic studies is the immortalised murine C₂C₁₂ cell line (Burattini et al. 2004).
99 The myoblast-derived C₂C₁₂ line displays cycles of 'active' to 'resting' that closely
100 resemble those of satellite cells that cycle between activated and quiescent states

Varshney et al.

101 (Yoshida et al. 1998). In contrast, primary muscle cell lines are directly established
102 from mouse tissue or human muscle biopsies (Zacharewicz et al. 2020). Although
103 more complex to establish and expensive to maintain, these provide an alternative
104 model that may be molecularly and metabolically more relevant to the human
105 physiology (Abdelmoez et al. 2020).

106 Here, we sought to understand the extent to which APA is associated with human
107 myoblast differentiation. We compared the transcriptomic and APA profiles of
108 differentiating primary human skeletal myoblasts (*pHSkMs*) with public data from
109 C₂C₁₂ myotubes (Abdelmoez et al. 2020). In contrast to our hypothesis, and unlike
110 what was observed in C₂C₁₂ cells, little APA and only a subtle trend toward 3' UTR
111 lengthening was found in differentiating *pHSkM*. Differences in proliferative capacity,
112 and stable expression of CPA factors, when compared to C₂C₁₂ cell differentiation,
113 where their expression is globally repressed, may explain the lack of APA in
114 differentiated *pHSkM*. In addition, the models displayed considerable differences in
115 metabolic gene ontology (GO) pathway enrichment, for example, a significant
116 induction of cholesterol biosynthesis in *pHSkM* cells was not recapitulated in C₂C₁₂
117 myotubes. Collectively, our data suggest caution should be exercised when using
118 immortalized models to infer human skeletal muscle biology and that the connection
119 between APA and differentiation might be more context-dependent than previously
120 thought.

Varshney et al.

121 **Results**

122 To investigate APA in human skeletal muscle differentiation, *pHSkM* from two
123 healthy donors, male (22.9 years old) and female (26.6 years old), were established,
124 passaged, and differentiated according to our previously published method (Lamon
125 et al. 2014; Zacharewicz et al. 2020). Samples were collected from the male-derived
126 HSM1 and female-derived HSM2 *pHSkM* lines, grown in replicates and harvested at
127 indicated time points (Fig. 1A). For comparison, the equivalent differentiation
128 regimen of C₂C₁₂ cells, as reported by Wang et al. (2019), is also provided. Peak
129 differentiation was achieved after seven days, as evidenced by extensive structural
130 rearrangement and the formation of polynucleated myotubes (Fig. 1B). RNA was
131 isolated (n = 24) for 3' end-focused RNA-seq library preparation (Quant-seq™) and
132 single-end sequencing on the Illumina HiSeq 3000 platform. Raw sequencing data
133 were processed using in-house bioinformatics pipelines for the statistical analysis of
134 gene expression and alternative polyadenylation (Harrison et al. 2015).
135 Multidimensional scaling (MDS) demonstrated strong reproducibility between
136 replicates and between donors (HSM1/2) along a shared trajectory of differentiation
137 (Fig. 1C). As expected, *pHSkMs* and differentiated myotubes displayed distinct gene
138 expression profiles, with the two *pHSkM* lines showing a high degree of replication
139 despite intrinsic genetic and sex differences between the donors (Fig. 1D). The
140 effectiveness of the differentiation protocol was supported by the downregulation of
141 cell-cycle genes (e.g., *ID1*, *ID3*, *CDC20*, and *PLK1*, indicated in orange), the
142 upregulation of key muscle-specific differentiation markers (e.g., *KLHL41* and
143 *MYOG*, green), and genes associated with skeletal muscle contraction (e.g., *MYH3*,
144 *MYH8*, *TNNC1*, *TNNC2*, and *TNNI1*, pink). Normalized data is available for
145 interactive search and visualization here

Varshney et al.

146 (<https://degust.erc.monash.edu/degust/compare.html?code=fcdeaa05fc9e7e33af29b804fec86d37#>

147) or as a list of all significantly regulated genes (Supplemental Table
148 S1).

149 Of the 12,518 detected genes, only 295 genes (2.6%) were significantly differentially
150 expressed (FDR < 0.05; Log2FC > 1) between the two donors on day 7 of
151 differentiation (Supplemental Table S2). A search for enriched GO terms failed to
152 identify any associated functional categories or obvious sex-linked differences.

153 Therefore, although further research into donor-specific differences could be of
154 interest, the overall replication between donors here was used to justify their use as
155 biological replicates in downstream analyses, for a total of four replicates at each
156 time point.

157 ***Primary Human skeletal muscle myoblasts undergo minor 3' UTR lengthening.***

158 Previously, Wang et al. (2019) reported global lengthening of 3' UTRs when
159 comparing undifferentiated C₂C₁₂ myoblasts (D0) and after differentiation (D4) using
160 the 3' READS method (Wang et al. 2019). The C₂C₁₂ time points reflected 95%
161 confluent myoblasts prior to induction of differentiation and differentiated myotubes,
162 respectively (Fig. 1A). analogous comparison in *pHSkM* (D0 vs. D7). For a direct
163 comparison and to minimize confounding effects from bioinformatic processing, the
164 raw data from Wang et al. (2019) were reanalysed using 'tail-tools'. Unlike what was
165 observed in C₂C₁₂ cells, very little APA was seen when we compared our *pHSkM*
166 data with previously published data. The only significantly shifted gene was *TPM2*,
167 for which the apparent shift was due to elevated expression of an alternative terminal
168 exon isoform (Fig. 2A). This represents an upregulation of an alternatively spliced
169 isoform, rather than 3' UTR switching *per se*, was confirmed by 3' RACE (Fig. S1A).

Varshney et al.

170 Because the RNA-seq methods used for both *pHSkM* and C₂C₁₂ data collection were
171 3'-focused, a side-by-side comparison was possible. Using raw data from Wang et
172 al. (2019) reprocessed with tail tools (Harrison et al. 2015), a shift score for each
173 gene in each sample was developed and implemented using the weighted matrix
174 (weitrix) package (Harrison 2022). A rank-based score (between -1 and 1) was used
175 to estimate the degree of APA between the groups of samples (myoblasts vs.
176 myotubes). A “confect” score with a specified False Discovery Rate of 0.05 was used
177 to rank genes by the magnitude of the APA shift (Harrison et al. 2019). To improve
178 the chance of APA detection through increased statistical power, undifferentiated
179 *pHSkMs* (D-1/D0) and differentiated myotubes (D1/D3/D5/D7) were pooled. Using
180 this approach, 195 genes displayed 3' UTR lengthening (positive shift) and 47
181 underwent 3' UTR shortening upon *pHSkM* differentiation (Fig. 2B; Supplemental
182 Table S3). In comparison, almost twice as many genes showed 3' UTR lengthening
183 (n = 382) or shortening (n = 114) in C₂C₁₂ cells (Fig. 2C; Supplemental Table S4).
184 Little correlation between gene expression change and APA was observed in either
185 model, suggesting that changes in 3' UTR usage occur independently of overall
186 transcript levels (Supplemental Fig. S1B-C).

187 It is possible that subtle differences in methodology affect APA detection. Therefore,
188 to quantify the global APA in an unbiased manner between unrelated experiments
189 beyond the number of genes affected, a new method was developed. By the
190 estimate of the sum of squares (ESS) approach, a numerical descriptor that
191 considers the number of APA events, the degree of shift, and its variance can be
192 calculated (see Methods). The ESS for APA in *pHSkM* was lower than that for C₂C₁₂
193 (ESS = 13.55 and 30.40, respectively), indicating that APA was significantly more
194 pronounced in the murine model. By way of comparison, the knockdown of Pcf11, a

Varshney et al.

195 core factor within the CPA with known impact on 3' UTR choice (Sadowski et al.
196 2003; Ogorodnikov et al. 2018; Wang et al. 2019; Turner et al. 2020), displayed a
197 high level of APA C₂C₁₂ cell line (ESS = 51.64) by re-analysis of (Wang et al. 2019)
198 data (Supplemental Fig. S1D). Thus, using siPcf11 as a positive control, *pHSkM* and
199 C₂C₁₂, underwent ~26% and ~59% APA, respectively.
200 To determine if the reduced APA in *pHSkM* was simply a less pronounced effect of
201 the same genes, the human homologues of mouse APA genes were retrieved using
202 Ensembl BioMart (38). Of the shared genes (n = 796) only 31 showed evidence of
203 APA in both models (Fig. 2D). Of these, most shifts were in the same direction (Fig.
204 2E). However, the overall correlation was low ($r^2 = 0.0039$), and clustering of the
205 confect values close to the y-axis indicates that for most APA events, neither the
206 genes nor their degree of shift was well conserved between *pHSkM* and C₂C₁₂
207 myotubes.
208 Previous studies have linked the differential expression of CPA factors to
209 proliferative potential (Gruber et al. 2014). Highly proliferative cells, such as cancer
210 cells, tend toward shorter 3' UTRs (Sandberg et al. 2008) such that the 3' UTR
211 length is inversely proportional to the level of available CPA in most cell types where
212 it has been measured (Sommerkamp et al. 2021; Turner et al. 2021). To determine
213 whether the observed difference between APA in the two myotube models could be
214 linked to the available CPA machinery, the expression of 33 previously identified
215 CPA genes (Gruber et al. 2014) was analysed. In contrast to C₂C₁₂ cells, where most
216 of the factors were differentially expressed, few of these showed significant
217 differential expression in *pHSkM* (Fig. 2F-G). Collectively, these data support a
218 model in which the paucity of APA associated with *pHSkM* differentiation compared

Varshney et al.

219 to that observed in the immortalised C₂C₁₂ cells might be linked to the stability of
220 CPA expression.

221 **Differentiating pHSkM and C₂C₁₂ myoblasts display major differences in gene
222 expression.**

223 To understand if the immortalised murine C₂C₁₂ cell recapitulated the human
224 differentiation processes more generally, we compared the gene expression profiles
225 of the two models. The normalised gene expression data for C₂C₁₂ differentiation are
226 available for interactive search here

227 (<https://degust.erc.monash.edu/degust/compare.html?code=c5d8df8af1b1dfe97477c254b30e876e#/>) or as a list of all significantly regulated genes (Supplemental Table
228 S5). A comparable pHSkM dataset included undifferentiated myoblasts (D0) and
229 differentiated myotubes (D7). In total, 12,518 human genes were detected, 1,052
230 (8.5%) of which were significantly upregulated and 984 (7.8%) were significantly
231 downregulated during differentiation (FDR < 0.05; Log2FC > 1) (Fig. 3A). In contrast,
232 of 11,799 detected murine genes, 1,932 (16.4%) were significantly upregulated and
233 1,710 (14.5%) were significantly downregulated (FDR < 0.05; Log2FC > 1) (Fig. 3B).
234 Interestingly, even with the statistical advantage of having twice the number of
235 replicates (n = 4 replicates from 2 independent donors), pHSkM showed fewer
236 differentially expressed genes and a lower magnitude of change than immortalised
237 C₂C₁₂ cells (n = 2).

238 To test for overlap between differentially expressed genes in the two models, human
239 orthologs of the significant murine genes were retrieved using Ensembl BioMart
240 (Smedley et al. 2009). Surprisingly, the overlap was modest. Of the 3,386 murine
241 genes with shared human annotations, only 796 genes were differentially regulated
242 in both datasets (Fig. 2C). To determine whether the lack of systematic replication

Varshney et al.

244 between the two models had implications for biological function, GO associations
245 were retrieved from the common and unique genes from the human or mouse lists
246 using Enrichr (Chen et al. 2013; Kuleshov et al. 2016). The common subset was
247 enriched, for processes specifically associated with muscle development, such as
248 muscle contraction and actin-myosin sliding, as well as genes related to DNA
249 processing, replication, and metabolic processes as expected, (Fig. 3D). Of the
250 human-specific differentially expressed genes (n = 1,240), pathways related to
251 alcohol, sterol, and cholesterol synthesis were enriched. A deeper analysis of the
252 enzymes in the cholesterol biosynthetic pathway showed that most genes were
253 upregulated in *pHSkM* (Fig. 3E). The complete gene set within the cholesterol
254 biosynthesis GO term (GO:0006695) is shown (Supplemental Fig. S2A). This was
255 not observed in mouse-specific differentially expressed genes (n = 2,590), where key
256 genes in the cholesterol biosynthetic pathway were significantly downregulated. In
257 contrast, differentiation in C₂C₁₂ cells was associated with the enrichment of genes
258 involved in the GO term, regulation of glycolytic processes (GO:0006110) (Fig. 3D,
259 Supplemental Fig. S2B).

260 Metabolic differences between myogenic models were recently highlighted in a
261 meta-analysis comparing differences in microarray-based gene expression data for
262 human, mouse, and rat myotubes, and corresponding tissue samples (Abdelmoez et
263 al. 2020). Although not specifically noted in their analysis, the expression of
264 *HMGCR*, the rate-limiting enzyme for cholesterol synthesis showed absolute
265 expression difference between models when visualised using their convenient tool
266 for interactive search <https://nicopillon.com/tools/muscle-models-profiling/>. Testing
267 for statistical significance of normalised microarray-based expression from their raw
268 data revealed that *HMGCR* and *DHCR24* were indeed expressed at a higher level in

Varshney et al.

269 *pHSkM* than C₂C₁₂ myotubes, and more importantly, also in their respective primary
270 tissue samples, indicating that the differences observed here were physiologically
271 relevant (Supplemental Fig. S3; Supplemental Table S6).

272 In total, 130 orthologs were significantly regulated in the opposite direction between
273 the models (Fig. 3F, Supplemental Table S7). Many such genes were supported by
274 external data including, but not exclusive to processes such as cytokine signalling
275 (SOCS2, SOCS3 and CCL2), small molecule transport (SLC43A3), glycolysis
276 (GPD1), lipogenesis (PCSK9 and FASN) (Supplemental Fig. S3). Collectively, these
277 results highlight that key metabolic pathways are differentially regulated between the
278 *pHSkM* and C₂C₁₂ during differentiation.

279 **Discussion**

280 An extensive literature links proliferation, exit from cell cycle, and differentiation to
281 APA [reviewed by Manley (2017)]. Accordingly, differentiation in the immortalised
282 C₂C₁₂ cell line, which is widely used to model muscle growth, function, and
283 metabolism, has been linked to the widespread 3' UTR lengthening (Li et al. 2015;
284 Wang et al. 2019). Here we sought to study APA in Primary human skeletal
285 myoblasts (*pHSkM*) differentiation, using established primary muscle cell lines
286 derived from donor muscle biopsies (Zacharewicz et al. 2020). In this study, we
287 show that unlike what has been observed in C₂C₁₂ myoblasts, differentiating *pHSkM*
288 displayed only minor APA after differentiation. Moreover, transcriptomic analysis
289 comparing *pHSkM* and C₂C₁₂ over a time course of differentiation revealed a
290 substantial lack of conservation in transcriptional profiles as well.
291 For ease of reuse, transcriptomic changes in *pHSkM* (Fig. 1) and C₂C₁₂
292 differentiation (Fig. 3) can be accessed via an interactive data portal
293 (<https://degust.erc.monash.edu/degust/compare.html?code=fcdeaa05fc9e7e33af29b>

Varshney et al.

294 [804fec86d37#](#) and
295 <https://degust.erc.monash.edu/degust/compare.html?code=c5d8df8af1b1dfe97477c254b30e876e#/> respectively). The differences in enriched GO pathways between the
296 two models were related to the major metabolic pathways (Fig. 3). In a meta-analysis
297 of publicly available microarray data, Abdelmoez et al. (2020) reported that
298 differentiated *pHSkM* and C₂C₁₂ myotubes employed different strategies for glucose
299 metabolism (Abdelmoez et al. 2020). While both *in vitro* models favoured glycolysis,
300 transcript levels suggest that C₂C₁₂ myotubes complete the oxidation of glucose
301 through mitochondrial respiration, whereas *pHSkM* myotubes appear to favour a
302 Warburg-like conversion of pyruvate to lactate. While both studies highlight
303 metabolic differences, the Abdelmoez study and our results were derived from a
304 fundamentally different experimental design. Abdelmoez et al. (2020), reported end-
305 point differences in mRNA amount (normalised signal strength) in myotubes of
306 different models as analysed by microarray (Abdelmoez et al. 2020). We report
307 internal gene expression changes (myoblasts vs. myotubes) and subsequently
308 compared the changes associated with differentiation between models.
309 Here, GO terms for carbohydrate metabolism were enriched in C₂C₁₂ compared to
310 the *pHSkM* model, confirming that C₂C₁₂ more closely recapitulates the metabolic
311 properties of fast muscle fibres. These glycolytic pathways were not enriched in
312 differentiating *pHSkM*, suggesting they may display mixed metabolic properties.
313 While C₂C₁₂ were originally established from mouse quadriceps muscle (Yaffe and
314 Sarel 1977), whose white portion comprises 100% of type IIb muscle fibres
315 (Campbell et al. 2001), *pHSkMs* were derived from skeletal muscle tissue samples
316 biopsied from the *Vastus Lateralis* muscle, which includes a mix of oxidative and
317 glycolytic fibres and up to 40% type I muscle fibres (Staron et al. 2000), with few sex-

Varshney et al.

319 specific differences. Surprisingly, we observed a robust upregulation of alcohol,
320 cholesterol, sterol, and steroid biosynthetic processes in *pHSkM* which were either
321 unaffected or downregulated in C₂C₁₂ myotubes (Fig. 3E, Supplemental Fig. S3).
322 Cholesterol is essential for myotube function, influencing electrical conduction,
323 signalling, and membrane fluidity via raft-like structures in the T-tubule (Barrientos et
324 al. 2017).
325 Cells are influenced by their genotype, including the complement of sex
326 chromosomes present (Zore et al. 2018). Despite being derived from donors of each
327 sex, the influence of sex chromosomes on gene expression in our *pHSkM* lines *in*
328 *vitro* was limited. While muscle-derived cells retain some epigenomic memory in
329 culture (Turner et al. 2020), the extent to which *pHSkM* lines maintain their identity,
330 outside of the context of normal physiological control, needs further examination.
331 The controlled environment of *pHSkM* cells in culture could reveal aspects of inter-
332 individual and sex-specific muscle metabolism that are masked by the direct analysis
333 of donor biopsies. However, such research would require a large number of donors
334 to achieve statistical power. Collectively, our data highlight significant gene
335 regulatory differences between the most common models used for *in vitro* study of
336 muscle biology. Based on these differences we caution against their interchangeable
337 use and suggest that, where it is feasible, the *pHSkM* model might more reliably
338 model human muscle physiology.
339 The relatively minor APA associated with differentiation in *pHSkM* *versus* C₂C₁₂ cells
340 was surprising. Multiple previous studies have connected myogenic differentiation to
341 APA. Indeed, the expression of APA molecular drivers and the switch between
342 proliferative and differentiated states has been well established. A simple
343 explanation for the difference in APA is that unlike C₂C₁₂ cells, which display marked

Varshney et al.

344 changes in CPA expression, *pHSkM* differentiation does not significantly alter their
345 expression levels (Fig. 2F). However, APA and the expression of CPA factors have
346 also been associated with proliferative potential (Gruber et al. 2014). Supporting a
347 role for proliferative rate in contributing to differences in APA reported here,
348 Abdelmoez et al. (2020), showed that C₂C₁₂ cells display as much as a 5-fold higher
349 proliferative capacity than *pHSkMs* (Abdelmoez et al. 2020). However, in addition to
350 legitimate differences in gene regulatory programs in the two models, we cannot rule
351 out that contributions by experimental factors such as the number of proliferative
352 cells remaining post-differentiation, and the transcriptomic contributions by any co-
353 cultured non-myocyte cells such as fibroblasts in *pHSkM* are likely to impact our
354 comparative analyses.

355 Finally, our data suggest that APA might be less marked in primary cell
356 differentiation than has been reported by study of transformed cells. Recent work
357 tracing APA in single-cell data (Agarwal et al. 2021; Wang et al. 2022; Zhou et al.
358 2022) brings a more fine-grained view of APA in native tissue than was previously
359 possible from bulk RNA-seq analyses of tissues or cells in culture. For example, by
360 extracting APA from single cells in cell-cycle resolved analyses (Wang et al. 2022),
361 in a cell type-specific manner in the brain (Yang et al. 2021) and across murine
362 during development, where a significant proportion (62%) of APA genes show longer
363 3'UTRs many (38%) do not (Agarwal et al. 2021). Furthermore, while some cell
364 types, such as neuronal lineages show pronounced evidence of 3'UTR lengthening
365 across developmental stages, others including myocytes do not (Agarwal et al.
366 2021). Thus, we believe that our findings fit into a broader narrative in which
367 identification, and interpretation of the regulation and function APA will become more
368 nuanced.

Varshney et al.

369 **Materials and Methods**

370 ***Primary human skeletal myoblast differentiation***

371 The *pHSkMs* were obtained from two healthy human donors and established as
372 previously reported (Zacharewicz et al. 2020) with ethical approval from the Deakin
373 University Human Research Ethics Committee (2018-388). Isolated *pHSkMs* from
374 the two donors, HSM1 (a 22.9 years old Caucasian male; 169 cm in height with a
375 body mass of 63 kg) and HSM2 (a 26.5 years old Caucasian female; 176 cm in
376 height with a body mass of 82.1 kg), were seeded in biological triplicates at 100,000
377 cells per well in 6-well plates with 2 ml of media. Cells were maintained and
378 expanded in proliferation media, Ham's F-10, with 20% Fetal Bovine Serum (FBS)
379 and 2.5 ng/mL (0.01%) basic Fibroblast Growth Factor (bFGF). Once cells reached
380 80% confluence, the proliferation medium was replaced with high glucose
381 Dulbecco's Modified Eagle's medium (DMEM) supplemented with 2% Horse Serum
382 to initiate differentiation. All media were supplemented with 0.5% penicillin-
383 streptomycin and 0.5% amphotericin B and refreshed every 48 h. The cells were
384 maintained at 37°C and 5% CO₂ in humid air. Brightfield microscope images were
385 taken of the cells at each stage of differentiation at 10x magnification at a scale of
386 1.525 pixel/µm (TS100 DS-Fi2-L3, Nikon). Images were converted to grayscale
387 without altering intensity or contrast.

388 ***RNA extraction and sequencing***

389 At each time point, two technical replicates were analysed per *pHSkM* line. For RNA
390 sample collection, the medium was aspirated and 600 µl TRIzol™ was added directly
391 to each 6-well plate well. After approximately 10 min of incubation at room
392 temperature, a cell scraper was used to ensure that all cellular material was
393 suspended. TRIzol™ extracts were stored at -80 °C until all samples were collected

Varshney et al.

394 for parallel processing. To isolate RNA, samples were thawed on ice, and the RNA
395 Directzol kit (Zymo) was used as per the manufacturer's protocol.

396 Quant-seq, a 3' end-focused RNA sequencing technique, was used to analyse global
397 gene expression and alternative poly(A) site selection in *pHSkM*. The libraries were
398 constructed using the QuantSeq 3' mRNA-Seq Library Prep Kit FWD for Illumina
399 (Lexogen). The libraries were prepared using 150 µg of input RNA according to the
400 manufacturer's instructions and sequenced at the Gandel Charitable Trust
401 Sequencing Centre (Australia) using 150bp single-end reads on a single lane of the
402 Illumina HiSeq 3000 instrument. The raw *pHSkM* data can be accessed from
403 [GSE168897](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE168897) (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE168897>). The
404 3' end-focused 3'READS RNA sequencing data were retrieved from

405 [GEO:GSE115232](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE115232)

406 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE115232>) (17): C₂C₁₂
407 myoblast (GSM3190450, GSM3190451), C₂C₁₂ myotube (GSM3190452,
408 GSM3190453), C₂C₁₂ siCtrl (GSM3171723, GSM3171724) and C₂C₁₂ siPcf11
409 (GSM3171727, GSM3171728).

410 **RNA-seq analysis**

411 The sequencing data was aligned to the human genome (GRCh38, Ensembl release
412 93) or mouse genome (GRCm38, Ensembl release 93) using STAR aligner (Dobin et
413 al. 2013), as implemented by the tail-tools pipeline (Harrison et al. 2015). To
414 visualize the read coverage, we used the Broad Institute's Integrative Genomics
415 Viewer (IGV) (Robinson et al. 2011).

416 Following alignment, the differential gene expression analysis was conducted using
417 Degust (Powell et al. 2019) using the Voom/Limma method (Law et al. 2014). To
418 eliminate genes with low expression, we used a minimum gene read count cut-off of

Varshney et al.

419 50 in at least one replicate, along with a minimum gene CPM cut-off of two in at least
420 three *pHSkM* or two *C₂C₁₂* replicates. Genes exhibiting a fold change greater than
421 two and a false discovery rate (FDR) < 0.05 were identified as statistically significant
422 and considered to be differentially expressed. High-dimensionality reduction was
423 achieved through Multidimensional Scaling (MDS) analysis using the
424 `limma::plotMDS()` function (Ritchie et al. 2015) with the top 2,500 differentially
425 expressed genes (DEGs).

426 In order to compare cross-species transcriptomic profiles, human orthologs of mouse
427 genes were obtained from Ensembl Biomart (release 93) (Smedley et al. 2009). To
428 identify genes with similar expression patterns, we executed hierarchical clustering
429 on z-scores derived from gene CPM values. The clustering was conducted with the
430 gplots package (version 3.1.0) (Warnes et al. 2020) using Euclidean distance
431 measures and a single linkage approach. We used Enrichr for Gene Ontology
432 analysis (Chen et al. 2013; Kuleshov et al. 2016) to identify enriched Biological
433 Process terms associated with the gene sets.

434 To further reinforce the gene expression analysis, we examined the normalised gene
435 counts from microarray data of human and mouse myotubes, along with
436 corresponding adult skeletal muscle tissue. This data is accessible on GitHub at
437 https://github.com/NicoPillon/Muscle_Models_Profiling/ (Abdelmoez et al. 2020). As
438 the data was non-normally distributed for the genes considered in this study, we
439 used the Kruskal-Wallis test with pairwise Dunn's multiple comparisons to identify
440 statistically significant differences in gene expression between species.

441 ***Differential APA analysis using weitrix***

442 Most APA detection methods do not control for variance well. Since samples with
443 more reads provide better estimates of APA, these can be accounted for using

Varshney et al.

444 observation-level weights similar to the limma-voom method (Law et al. 2014) for
445 gene expression data. Here, a new method that improves on our previous APA
446 detection (Turner et al. 2021) was implemented using the weitrix R package
447 (Harrison 2022). For *pHSkM*, the expression of proliferating samples (D-1/D0, n = 8)
448 and differentiated samples (D1/D3/D5/D7, n = 16) were compared. Similarly, for
449 *C₂C₁₂* the expression of proliferating samples (D0, n = 2) and differentiated samples
450 (D4, n = 2) were compared. Meanwhile, for *C₂C₁₂* siPcf11, the comparison focused
451 on the Pcf11 knock-down (siPcf11, n = 2) and control (siCtrl, n = 2).
452 Specifically, we calculated summary scores of the APA state of each gene in each
453 sample, compared to an average over all samples, with values between -1 (proximal)
454 and 1 (distal). From these per-sample scores, we can estimate the change in APA
455 state between the groups of samples. We also determine whether any such change
456 is statistically significant and place an inner confidence bound on the change using
457 the Topconfects method (Harrison et al. 2019). This method can be used for genes
458 with two or more APA sites.

459 For each gene i , a set of $n_{site\ i}$ APA sites were identified. Let the read count in
460 sample j at site k be $c_{i,j,k}$, the total count across samples be $c_{i,\cdot,j}$, across sites be
461 $c_{i,j,\cdot}$ and across samples and sites be $c_{i,\cdot,\cdot}$. Assume that sites have been ordered from
462 the furthest upstrand to the furthest downstrand. The rank-based score for all the
463 reads at a site, $s_{site\ i,k}$, was calculated using the proportion of reads upstrand of the
464 site minus the proportion of reads downstrand of the site.

$$465 \quad s_{site\ i,k} = \frac{1}{c_{i,\cdot,\cdot}} \sum_{k=1}^{n_{site\ i}} \left(\sum_{l=1}^{k-1} c_{i,\cdot,l} - \sum_{l=k+1}^{n_{site\ i}} c_{i,\cdot,l} \right)$$

Varshney et al.

466 Then, for each gene i and sample j , the score is calculated as the average of the
467 site score of each read.

$$468 \quad s_{i,j} = \frac{1}{c_{i,j,\cdot}} \sum_{k=1}^{n_{site\ i}} s_{site\ i,k} c_{i,j,k}$$

469 The variance of the average of $c_{i,j,\cdot}$ such reads will be $1/c_{i,j,\cdot}$ times the read-level
470 variance. Different amounts of noise in particular observations are accounted for in
471 linear models using observation-level weights that are inversely proportional to the
472 variance. The appropriate weights are $c_{i,j,\cdot}$. If there were no reads for a particular
473 sample in a particular gene, then the weight was zero, indicating missing data. Using
474 these weights, an initial linear model was fitted to each gene.

475 If there is biological variation between individual samples, the variance will be
476 somewhat higher than the suggested initial weights. If there are real differences
477 between the groups of samples, the variance within each group will be somewhat
478 lower. Therefore, we sought to refine the initial weights. We can also potentially use
479 an estimate of the per-read variance for each gene under the assumption that the
480 reads were observed at random.

$$481 \quad \sigma_{read\ i}^2 = \frac{1}{c_{i,\cdot,\cdot}} \sum_{k=1}^{n_{site\ i}} s_{site\ i,k}^2 c_{i,\cdot,k}$$

482 Calibrated weights were determined using residuals from the initial linear models
483 fitted for each gene. A gamma GLM with a log link function of the squared residuals
484 was fitted in terms of the log per-read variance $\sigma_{read\ i}^2$ and a cubic spline curve of the
485 log number of reads. This model was fitted to the entire dataset. The inverse of the
486 predictions of this model provided a new set of weights. In other words, the new

Varshney et al.

487 weights are calculated as a calibrated function of the number of reads and per-read
488 variance.

489 Linear models were again fitted for each gene using the new weights. The final step
490 of calibration is provided by Empirical Bayes squeezing of the residual variance for
491 each gene, as in the limma. Shift effect sizes between groups are estimated as
492 contrasts of linear model coefficients, as in the limma. The calibrated weights
493 allowed accurate standard errors of these contrasts to be calculated.

494 **Quantifying APA with the Estimated Sum of Squares (ESS)**

495 To quantify the extent of APA between independent samples beyond the raw
496 number of genes, a new approach was required. A quantity that measures this is the
497 sum of the squared APA shifts for all the genes. However, estimated shifts contain
498 estimation errors that inflate the sum of squared shifts, and an adjustment for this is
499 necessary.

500 For each gene, i , we regarded the estimated shift as a random variable S_i composed
501 of a true shift t_i plus a random error ε_i .

502
$$S_i = t_i + \varepsilon_i$$

503 We seek an unbiased estimate of the sum of squares of the true shifts $\sum_i t_i^2$. We
504 assume $E(\varepsilon_i) = 0$, so $E(\varepsilon_i^2) = \sigma_i^2$ with σ_i^2 is the variance of the estimation error. The
505 differential APA test provides an estimate of this variance as part of its output V_i ,
506 $E(V_i) = \sigma_i^2$.

507
$$S_i^2 = t_i^2 + 2t_i\varepsilon_i + \varepsilon_i^2$$

508
$$\Rightarrow E(S_i^2) = t_i^2 + 2t_iE(\varepsilon_i) + E(\varepsilon_i^2)$$

509
$$= t_i^2 + E(V_i)$$

Varshney et al.

510
$$\Rightarrow t_i^2 = E(S_i^2 - V_i)$$

511 Thus, the square of the estimated shift minus the estimated variance is an unbiased
512 estimate of t_i^2 , and summing these estimates over all genes provides an unbiased
513 estimate of the sum of squares of the true shifts. Because this is adjusted for
514 measurement accuracy, values are comparable between experiments performed
515 with differing numbers of replicates, sequencing depth, and other sources of
516 variation.

517 **3' RACE validation**

518 The TVN-Poly(A) Test (TVN-PAT) was used to reverse transcribe polyadenylated
519 RNA into cDNA for PCR-based 3' RACE (Janicke et al. 2012). cDNA, prepared from
520 150 ng of total RNA input from each timepoint, was amplified with forward primers
521 designed upstream of the proximal or distal poly(A) cleavage site: proximal forward
522 primer (5'-GGAGCAGAAATTGCCAACAT-3') and distal forward primer (5'-
523 CGAGATTCAACCAGACCTTGG-3'), along with a universal reverse primer (PAT
524 reverse primer – 5'-GCGAGCTCCGCGGCCGCGTTTTTTTTTT-3'). For cDNA
525 amplification, 5 μ l of the cDNA, diluted at a ratio of 1:11, was PCR amplified using
526 AmpliTaq Gold 360 under the following conditions: 95°C for 10 min; 95°C for 20 s,
527 60°C for 20 s, 72°C for 30 s (27 cycles); and 72°C for 5 min. The amplified product
528 was separated by TBE electrophoresis with 2% ultrapure agarose including
529 SYBRsafe for imaging. Gels were imaged with Amersham™ Imager 600. Linear
530 uniform adjustments were made to image brightness and contrast.

531

Varshney et al.

532 **Data availability**

533 The sequencing data produced in this study have been submitted to GEO under the

534 accession number [GSE168897](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE168897)

535 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE168897>).

536 **Supplemental Material**

537 Supplemental Figures S1-S3

538 Supplemental Tables S1-S7

Varshney et al.

539 **Acknowledgements**

540 We thank the members of the Beilharz and Lamon labs for their constructive
541 feedback on the manuscript. The MHTP Gandel Medical Genomics Facility and
542 Monash Bioinformatics Platform are acknowledged for providing technical and
543 infrastructure support.

544 **Grants**

545 AV was supported by the Australian Postgraduate Research Award. SL was
546 supported by an ARC Future Fellowship (FT210100278). THB was supported by a
547 Monash Biomedicine Discovery Fellowship and grants from ARC (DP170100569 and
548 FT180100049).

549 **Disclosures**

550 No conflicts of interest are declared by the authors.

551 **Author contribution**

552 AV performed bioinformatics and RNA analyses, PFH developed the weitrix and
553 ESS bioinformatic packages for robust differential APA analysis, SEA isolated and
554 differentiated donor *pHSkMs*, and AS performed RNA-seq and guided wet-lab RNA
555 research. BD, SL, and THB conceived of and guided the study. AV, PFH, SL and
556 THB wrote the manuscript. All authors provided feedback on the manuscript.

557 **Conflict of Interest**

558 The authors declare no conflicts of interest.

Varshney et al.

559 **References**

560 Abdelmoez AM, Puig LS, Smith JAB, Gabriel BM, Savikj M, Dollet L, Chibalin AV,
561 Krook A, Zierath JR, Pillon NJ. 2020. Comparative profiling of skeletal muscle
562 models reveals heterogeneity of transcriptome and metabolism. *Am J Physiol
563 Cell Physiol* **318**: 615-626.

564 Agarwal V, Lopez-Darwin S, Kelley DR, Shendure J. 2021. The landscape of
565 alternative polyadenylation in single cells of the developing mouse embryo.
566 *Nat Commun* **12**: 5101.

567 Barrientos G, Sánchez-Aguilera P, Jaimovich E, Hidalgo C, Llanos P. 2017.
568 Membrane Cholesterol in Skeletal Muscle: A Novel Player in Excitation-
569 Contraction Coupling and Insulin Resistance. *Journal of diabetes research*
570 **2017**: 3941898-3941898.

571 Beilharz TH, See MM, Boag PR. 2019. 3'-UTRs and the Control of Protein
572 Expression in Space and Time. *Adv Exp Med Biol* **1203**: 133-148.

573 Bentzinger CF, Wang YX, Rudnicki MA. 2012. Building Muscle: Molecular
574 Regulation of Myogenesis. *Cold Spring Harbor perspectives in biology* **4**.

575 Burattini S, Ferri P, Battistelli M, Curci R, Luchetti F, Falcieri E. 2004. C2C12 murine
576 myoblasts as a model of skeletal muscle development: morpho-functional
577 characterization. *European Journal of Histochemistry* **48**: 223-234.

578 Campbell WG, Gordon SE, Carlson CJ, Pattison JS, Hamilton MT, Booth FW. 2001.
579 Differential global gene expression in red and white skeletal muscle. *Am J
580 Physiol Cell Physiol* **280**: C763-768.

581 Chal J, Pourquié O. 2017. Making muscle: Skeletal myogenesis in vivo and in vitro.
582 *Development (Cambridge)* **144**: 2104-2122.

583 Charge SBP, Rudnicki MA. 2004. Cellular and molecular regulation of muscle
584 regeneration. *Physiological reviews* **84**: 209-238.

585 Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, Clark NR, Ma'ayan A.
586 2013. Enrichr: Interactive and collaborative HTML5 gene list enrichment
587 analysis tool. *BMC Bioinformatics* **14**.

588 Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson
589 M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner.
590 *Bioinformatics (Oxford, England)* **29**: 15-21.

591 Giammartino DCD, Nishida K, Manley JL. 2011. Mechanisms and consequences of
592 alternative polyadenylation. *Molecular cell* **43**: 853-853.

593 Grassi E, Santoro R, Umbach A, Grosso A, Oliviero S, Neri F, Conti L, Ala U,
594 Provero P, Dicunto F et al. 2019. Choice of alternative polyadenylation sites,
595 mediated by the rna-binding protein Elavl3, plays a role in differentiation of
596 inhibitory neuronal progenitors. *Frontiers in Cellular Neuroscience* **12**: 1-15.

597 Gruber AR, Martin G, Keller W, Zavolan M. 2014. Means to an end: mechanisms of
598 alternative polyadenylation of messenger RNA precursors. *Wiley
599 Interdisciplinary Reviews: RNA* **5**: 183-196.

600 Harrison PF. 2022. weitrix: Tools for matrices with precision weights, test and
601 explore weighted or sparse data.

602 Harrison PF, Pattison AD, Powell DR, Beilharz TH. 2019. Topconfects: a package for
603 confident effect sizes in differential expression analysis provides a more
604 biologically useful ranked gene list. *Genome Biology* **20**: 1-12.

605 Harrison PF, Powell DR, Clancy JL, Preiss T, Boag PR, Traven A, Seemann T,
606 Beilharz TH. 2015. PAT-seq: A method to study the integration of 3'-UTR

Varshney et al.

607 dynamics with gene expression in the eukaryotic transcriptome. *Rna* **21**:
608 1502-1510.

609 Janicke A, Vancuylenberg J, Boag PR, Traven A, Beilharz TH. 2012. ePAT A simple
610 method to tag adenylated RNA to measure poly(A)tail length and other 3'
611 RACE applications.pdf. pp. 1289-1295.

612 Ji Z, Lee JY, Pan Z, Jiang B, Tian B. 2009. Progressive lengthening of 3'
613 untranslated regions of mRNAs by alternative polyadenylation during mouse
614 embryonic development. *Proceedings of the National Academy of Sciences of
the United States of America* **106**: 7028-7033.

615 Ji Z, Tian B. 2009. Reprogramming of 3' untranslated regions of mRNAs by
616 alternative polyadenylation in generation of pluripotent stem cells from
617 different cell types. *PLoS one* **4**: e8419.

618 Jiwlawat N, Lynch E, Jeffrey J, Van Dyke JM, Suzuki M. 2018. Current progress and
619 challenges for skeletal muscle differentiation from human pluripotent stem
620 cells using transgene-free approaches. *Stem Cells International* **2018**.

621 Kandhari N, Kraupner-Taylor CA, Harrison PF, Powell DR, Beilharz TH. 2021. The
622 Detection and Bioinformatic Analysis of Alternative 3' UTR Isoforms as
623 Potential Cancer Biomarkers. *International Journal of Molecular Sciences*
624 **2021, Vol 22, Page 5322** **22**: 5322-5322.

625 Korthuis RJ. 2011. Anatomy of Skeletal Muscle and Its Vascular Supply.

626 Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, Koplev S,
627 Jenkins SL, Jagodnik KM, Lachmann A et al. 2016. Enrichr: a comprehensive
628 gene set enrichment analysis web server 2016 update. *Nucleic acids research*
629 **44**: W90-W97.

630 Lackford B, Yao C, Charles GM, Weng L, Zheng X, Choi EA, Xie X, Wan J, Xing Y,
631 Freudenberg JM et al. 2014. Fip1 regulates mRNA alternative polyadenylation
632 to promote stem cell self-renewal. *EMBO Journal* **33**: 878-889.

633 Lamon S, Zacharewicz E, Stephens AN, Russell AP. 2014. EPO-receptor is present
634 in mouse C2C12 and human primary skeletal muscle cells but EPO does not
635 influence myogenesis. *Physiological Reports* **2**: 1-22.

636 Law CW, Chen Y, Shi W, Smyth GK. 2014. voom: precision weights unlock linear
637 model analysis tools for RNA-seq read counts. *Genome Biology* **2014** **15**:2
638 **15**: 1-17.

639 Li W, You B, Hoque M, Zheng D, Luo W, Ji Z, Park JY, Gunderson SI, Kalsotra A,
640 Manley JL et al. 2015. Systematic profiling of poly(A)+ transcripts modulated
641 by core 3' end processing and splicing factors reveals regulatory rules of
642 alternative cleavage and polyadenylation. *PLoS genetics* **11**: e1005166.

643 Manley BTaJL. 2017. Alternative polyadenylation of mRNA precursors Bin. *nature
reviews molecular cell biology* **18**: 18-30.

644 Masamha CP, Xia Z, Yang J, Albrecht TR, Li M, Shyu AB, Li W, Wagner EJ. 2014.
645 CFIm25 links alternative polyadenylation to glioblastoma tumour suppression.
646 *Nature* **510**: 412-416.

647 Mayr C, Bartel DP. 2009. Widespread Shortening of 3'UTRs by Alternative Cleavage
648 and Polyadenylation Activates Oncogenes in Cancer Cells. *Cell* **138**: 673-684.

649 Meng F, Zhao Y, Titus T, Zhang C, Postlethwait JH. 2018. Brain of the blind:
650 Transcriptomics of the golden-line cavefish brain. *Current Zoology* **64**: 765-
651 773.

652 Mitsche MA, McDonald JG, Hobbs HH, Cohen JC. 2015. Flux analysis of cholesterol
653 biosynthesis in vivo reveals multiple tissue and cell-type specific pathways.
654 *eLife* **4**: 1-21.

Varshney et al.

657 Nimura K, Yamamoto M, Takeichi M, Saga K, Takaoka K, Kawamura N, Nitta H,
658 Nagano H, Ishino S, Tanaka T et al. 2016. Regulation of alternative
659 polyadenylation by Nkx2-5 and Xrn2 during mouse heart development. *eLife*
660 5.

661 Ogorodnikov A, Levin M, Tattikota S, Tokalov S, Hoque M, Scherzinger D, Marini F,
662 Poetsch A, Binder H, Macher-Göppinger S et al. 2018. Transcriptome 3'end
663 organization by PCF11 links alternative polyadenylation to formation and
664 neuronal differentiation of neuroblastoma. *Nature Communications* 9: 1-16.

665 Pereira-Castro I, Moreira A. 2021. On the function and relevance of alternative 3'-
666 UTRs in gene expression regulation. *Wiley interdisciplinary reviews RNA* 12:
667 e1653.

668 Powell D, Milton M, Perry A, Santos K. 2019. drpowell/degust 4.1.1.

669 Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. 2015. Limma
670 powers differential expression analyses for RNA-sequencing and microarray
671 studies. *Nucleic Acids Research* 43: e47-e47.

672 Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G,
673 Mesirov JP. 2011. Integrative genomics viewer. *Nature biotechnology* 29: 24-
674 26.

675 Sadowski M, Dichtl B, Hubner W, Keller W. 2003. Independent functions of yeast
676 Pcf11p in pre-mRNA 3' end processing and in transcription termination. *The
677 EMBO journal* 22: 2167-2177.

678 Sandberg R, Neilson JR, Sarma A, Sharp PA, Burge CB. 2008. Proliferating cells
679 express mRNAs with shortened 3' untranslated regions and fewer microRNA
680 target sites. *Science* 320: 1643-1647.

681 Smedley D, Haider S, Ballester B, Holland R, London D, Thorisson G, Kasprzyk A.
682 2009. BioMart - Biological queries made easy. *BMC genomics* 10: 22-22.

683 Soetanto R, Hynes CJ, Patel HR, Humphreys DT, Evers M, Duan G, Parker BJ,
684 Archer SK, Clancy JL, Graham RM et al. 2016. Role of miRNAs and
685 alternative mRNA 3'-end cleavage and polyadenylation of their mRNA targets
686 in cardiomyocyte hypertrophy. *Biochimica et Biophysica Acta (BBA) - Gene
687 Regulatory Mechanisms* 1859: 744-756.

688 Sommerkamp P, Cabezas-Wallscheid N, Trumpp A. 2021. Alternative
689 Polyadenylation in Stem Cell Self-Renewal and Differentiation. *Trends Mol
690 Med* 27: 660-672.

691 Staron RS, Hagerman FC, Hikida RS, Murray TF, Hostler DP, Crill MT, Ragg KE,
692 Toma K. 2000. Fiber type composition of the vastus lateralis muscle of young
693 men and women. *J Histochem Cytochem* 48: 623-629.

694 Tian B, Manley JL. 2017. Alternative polyadenylation of mRNA precursors. *Nat Rev
695 Mol Cell Biol* 18: 18-30.

696 Turner DC, Gorski PP, Maasar MF, Seaborne RA, Baumert P, Brown AD, Kitchen
697 MO, Erskine RM, Dos-Remedios I, Voisin S et al. 2020. DNA methylation
698 across the genome in aged human skeletal muscle tissue and muscle-derived
699 cells: the role of HOX genes and physical activity. *Scientific reports* 10.

700 Turner RE, Harrison PF, Swaminathan A, Kraupner-Taylor CA, Goldie BJ, See M,
701 Peterson AL, Schittenhelm RB, Powell DR, Creek DJ et al. 2021. Genetic and
702 pharmacological evidence for kinetic competition between alternative poly(A)
703 sites in yeast. *Elife* 10.

704 Turner RE, Pattison AD, Beilharz TH. 2018. Alternative polyadenylation in the
705 regulation and dysregulation of gene expression. *Semin Cell Dev Biol* 75: 61-
706 69.

Varshney et al.

707 Wang J, Chen W, Yue W, Hou W, Rao F, Zhong H, Qi Y, Hong N, Ni T, Jin W. 2022.
708 Comprehensive mapping of alternative polyadenylation site usage and its
709 dynamics at single-cell resolution. *Proc Natl Acad Sci U S A* **119**:
710 e2113504119.

711 Wang R, Zheng D, Wei L, Ding Q, Tian B. 2019. Regulation of Intronic
712 Polyadenylation by PCF11 Impacts mRNA Expression of Long Genes. *Cell*
713 Reports **26**: 2766-2778.

714 Warnes GR, Bolker B, Bonebakker L, Gentleman R, Huber W, Liaw A, Lumley T,
715 Maechler M, Magnusson A, Moeller S et al. 2020. gplots: Various R
716 Programming Tools for Plotting Data [R package gplots version 3.1.0].

717 Yaffe D, Saxel O. 1977. Serial passaging and differentiation of myogenic cells
718 isolated from dystrophic mouse muscle. *Nature* **270**: 725-727.

719 Yang Y, Paul A, Bach TN, Huang ZJ, Zhang MQ. 2021. Single-cell alternative
720 polyadenylation analysis delineates GABAergic neuron types. *BMC Biol* **19**:
721 144.

722 Yang Y, Wu X, Yang W, Jin W, Wang D, Yang J, Jiang G, Zhang W, Niu X, Gong J.
723 2022. Dynamic alternative polyadenylation during iPSC differentiation into
724 cardiomyocytes. *Comput Struct Biotechnol J* **20**: 5859-5869.

725 Yoshida N, Yoshida S, Koishi K, Masuda K, Nabeshima Y. 1998. Cell heterogeneity
726 upon myogenic differentiation: down-regulation of MyoD and Myf-5 generates
727 'reserve cells'. *Journal of cell science* **111**: 769-779.

728 Yuan F, Hankey W, Wagner EJ, Li W, Wang Q. 2021. Alternative polyadenylation of
729 mRNA and its role in cancer. *Genes Dis* **8**: 61-72.

730 Zacharewicz E, Kalanon M, Murphy RM, Russell AP, Lamon S. 2020. MicroRNA-
731 99b-5p downregulates protein synthesis in human primary myotubes. *Am J*
732 *Physiol Cell Physiol* **319**: C432-C440.

733 Zhou R, Xiao X, He P, Zhao Y, Xu M, Zheng X, Yang R, Chen S, Zhou L, Zhang D et
734 al. 2022. SCAPE: a mixture model revealing single-cell polyadenylation
735 diversity and cellular dynamics during cell differentiation and reprogramming.
736 *Nucleic Acids Res* **50**: e66.

737 Zore T, Palafox M, Reue K. 2018. Sex differences in obesity, lipid metabolism, and
738 inflammation-A role for the sex chromosomes? *Molecular metabolism* **15**: 35-
739 44.

740

Varshney et al.

741 **Figure legends**

742 **FIGURE 1. Reproducible differentiation of *pHSkM* cell lines.**

743 **(A)** Schematic representation of primary Human Skeletal Myoblast (*pHSkM*)

744 differentiation (top) or C₂C₁₂ cells as reported by Wang et al. (2019) (bottom).

745 Sample collection was at the time points indicated.

746 **(B)** Grayscale brightfield microscopy images of *pHSkMs* at the indicated days of

747 differentiation. The scale bar for all images is 100 μ m.

748 **(C)** MDS analysis of male-derived HSM1 (circle) and female-derived HSM2 (triangle)

749 *pHSkMs*. The two replicates at each time point show discrete separation based on

750 the differentiation of the myoblast lines (dim 1). For this analysis, the log2-normalised

751 gene CPM were used for the top 2,500 genes.

752 **(D)** Hierarchical clustering analysis of the top 75 significantly differentially expressed

753 genes during *pHSkM* differentiation, illustrates overall reproducibility between the

754 male-derived HSM1 and female-derived HSM2 cell lines (n = 2 replicates at each

755 time point). Key gene categories, including downregulation of cell-cycle genes

756 (orange), and upregulation of skeletal muscle contraction (pink) and muscle-specific

757 differentiation markers (green) support efficacy of diff. The z-scores of gene CPM

758 were clustered using Euclidean distance and single linkage method. Red is

759 upregulated and blue is downregulated, as shown in the key.

760

761 **FIGURE 2. Differential CPA factor expression between models could explain**

762 **APA differences.**

763 **(A)** IGV representation of reads mapping to *TPM2* 3' UTR (Human version hg38).

764 Proliferating *pHSkM* (HSM2 D0 R1 & R2) have reads (red) mapping to the distal

Varshney et al.

765 poly(A) site of *TPM2*, while the primary myotubes (HSM2 D7 R1 & R2) have a
766 greater number of reads (blue) mapping to the proximal poly(A) site.
767 **(B-C)** Scatter plot of confect against average gene expression to visualise the shift in
768 proportion of proximal versus distal poly(A) site usage. ESS (Estimated Sum of
769 Squares) is an overall metric of the amount of APA, incorporating both the number of
770 genes involved and the magnitude of shifting, intended to be broadly comparable
771 across species and experiments (details in the Methods section). The degree of shift
772 is quantitated by the estimated sum of squares (ESS) value. **(B)** *pHSkM* underwent
773 less shift during differentiation ($n = 242$, ESS = 13.55) when compared to **(C)** *C₂C₁₂*
774 differentiation ($n = 496$, ESS = 31.42). Red represents 3' UTR lengthening and blue
775 represents 3' UTR shortening.
776 **(D)** Venn overlap of genes undergoing a shift in poly(A) site usage during *pHSkM* (D-
777 1/D0 vs. D1/D3/D5/D7) and *C₂C₁₂* (D0 vs. D4; $n = 2$ replicates at each time point)
778 differentiation, with 31 genes in common.
779 **(E)** Scatter plot of 31 overlapping shifting genes in *pHSkM* and *C₂C₁₂* muscle
780 differentiation.
781 **(F)** Negative log10-normalised FDR of cleavage and polyadenylation (CPA) factors
782 from (3). Myotubes vs. myoblasts, D0 vs. D7 in *pHSkM* and D0 vs. D4 in *C₂C₁₂*. FDR
783 > 0.05 in dark grey. Nova1 expression was not detected in *C₂C₁₂*.
784
785 **FIGURE 3. C₂C₁₂ and pHSkM differ on a molecular and metabolic level.**
786 **(A-B)** Volcano plot of statistical significance and fold change in proliferated versus
787 differentiated *pHSkM* (D0 vs. D7; $n = 4$ replicates from 2 independent donors at each
788 time point) **(A)** or *C₂C₁₂* (D0 vs. D4; $n = 2$ replicates at each time point) **(B)** cells. Red

Varshney et al.

789 indicates upregulated genes, and blue indicates downregulated genes (FDR < 0.05,
790 Log2FC > 1).

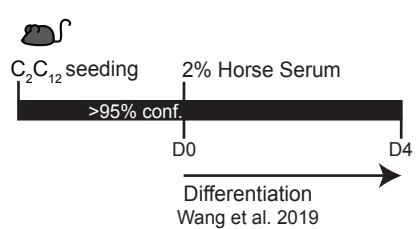
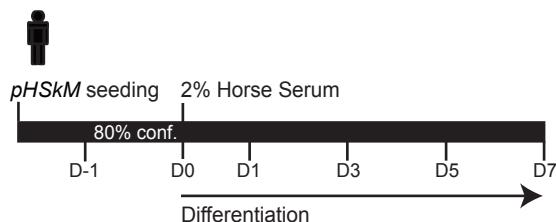
791 **(C)** Venn overlap of DEGs in *pHSkM* differentiation and human orthologs of DEGs in
792 C₂C₁₂ differentiation, with 796 genes common.

793 **(D)** Hierarchical clustering analysis of 796 overlapping genes during *pHSkM* and
794 C₂C₁₂ differentiation. The z-scores of CPM were clustered using Euclidean distance
795 and the single linkage method. Red signifies upregulated and blue downregulated,
796 as shown in the key.

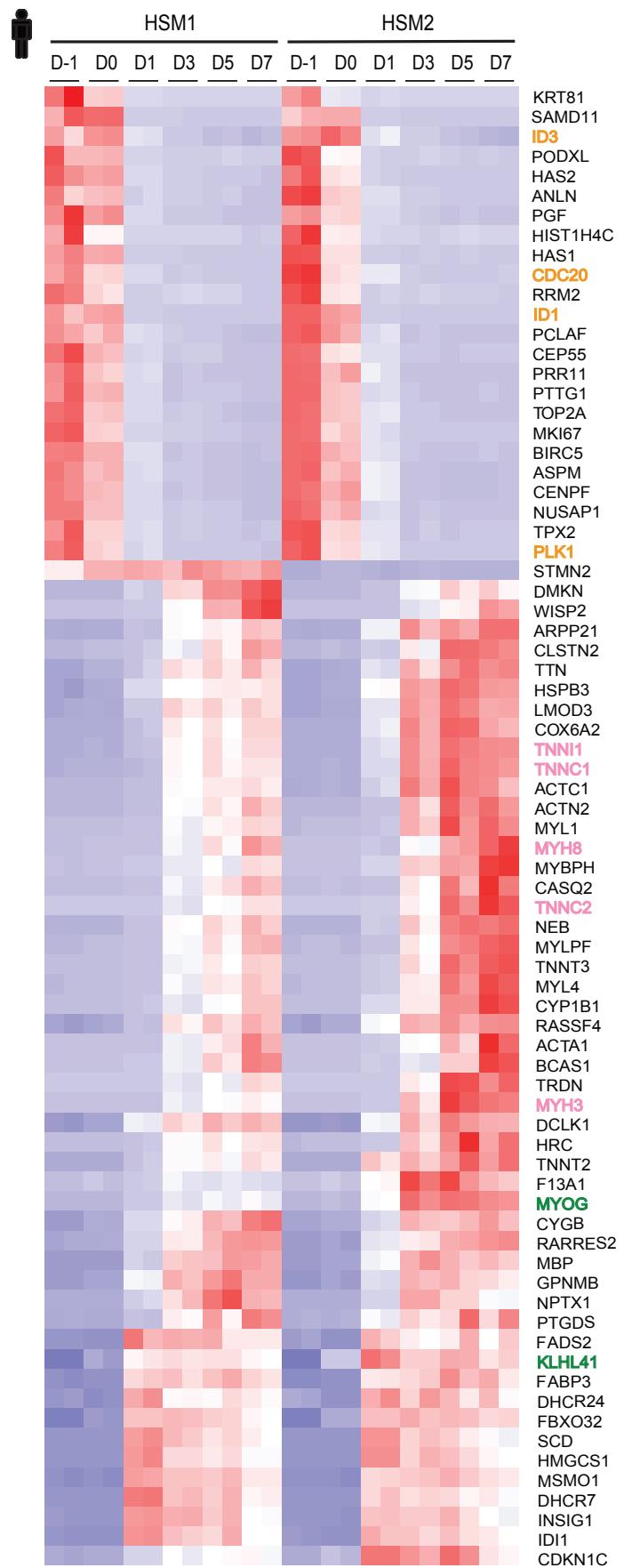
797 **(E)** Gene Ontology (GO) analysis of DEGs specific to *pHSkM* (n = 1,240), common
798 overlap (n = 796), and C₂C₁₂ (n = 2,590). The five most significant GO terms were
799 then selected for each subset. Size and colour correspond to the gene ratio and
800 adjusted p-value, respectively. With orange being the most significant.

801 **(F)** Genes involved in cholesterol biosynthesis pathways (Mitsche et al. 2015; Meng
802 et al. 2018). Red and blue indicate upregulated and downregulated genes,
803 respectively. Black indicates no significant changes in gene expression during
804 skeletal muscle differentiation.

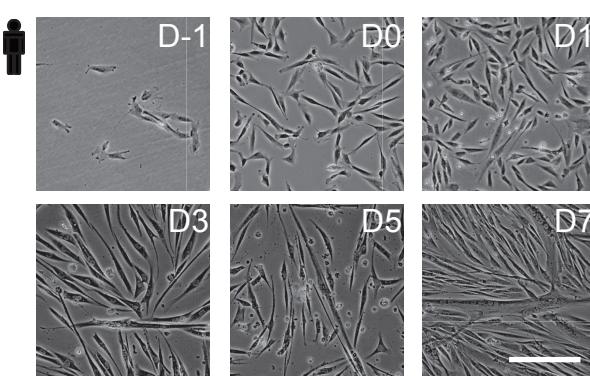
A



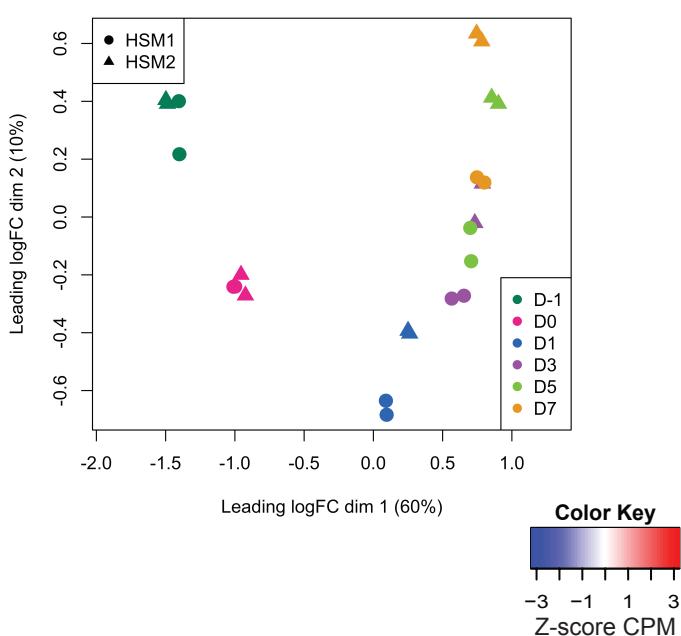
D



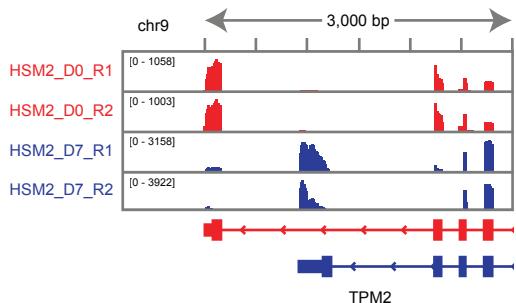
B



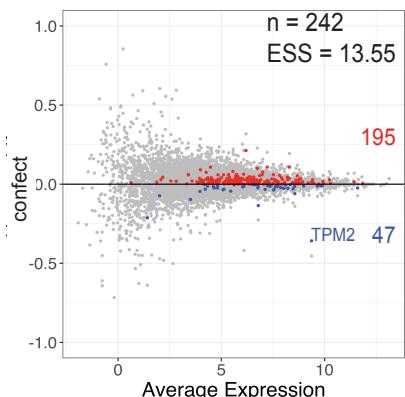
C



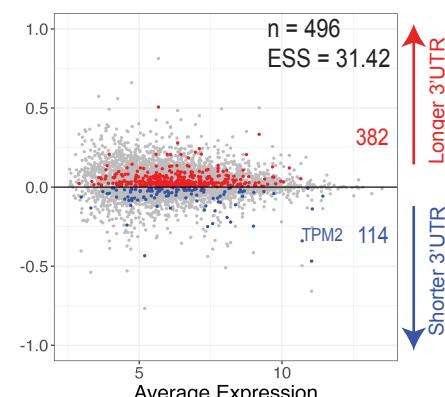
A



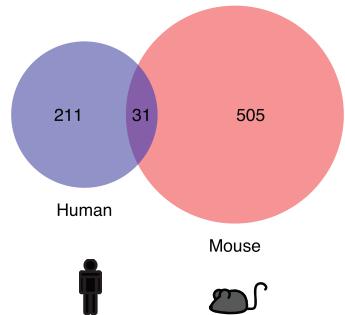
B



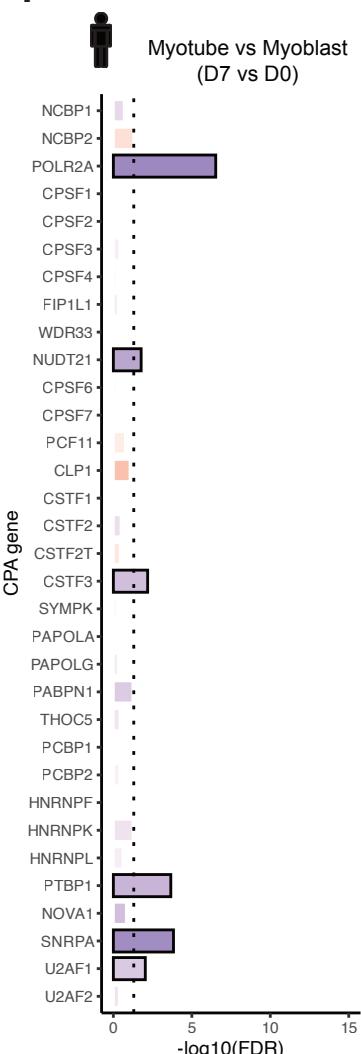
C



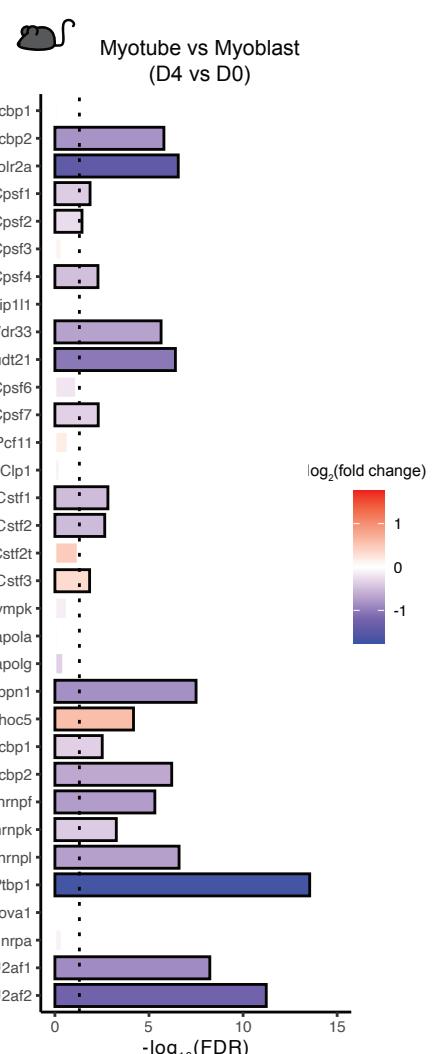
D



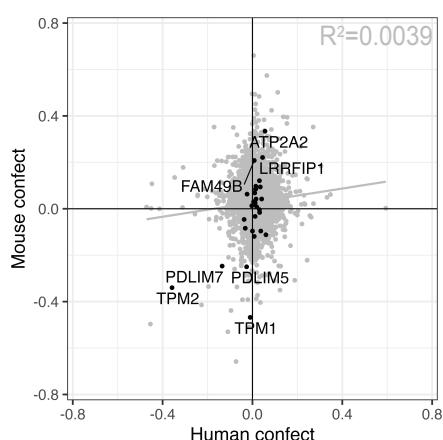
F



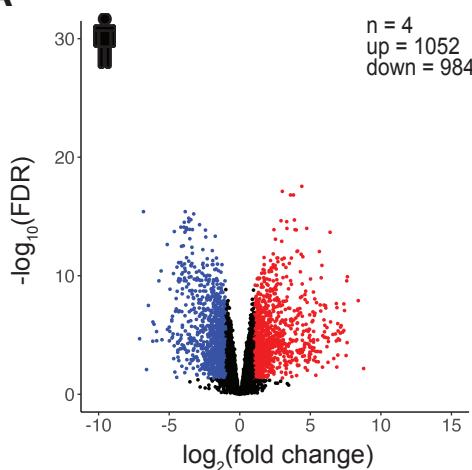
G



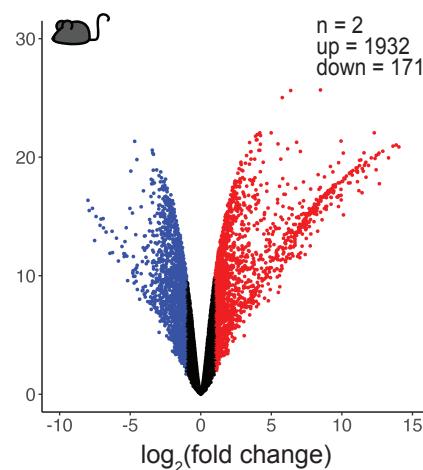
E



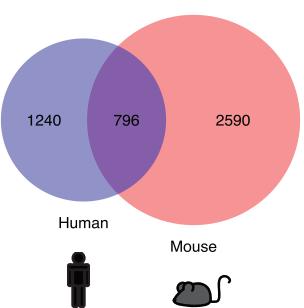
A



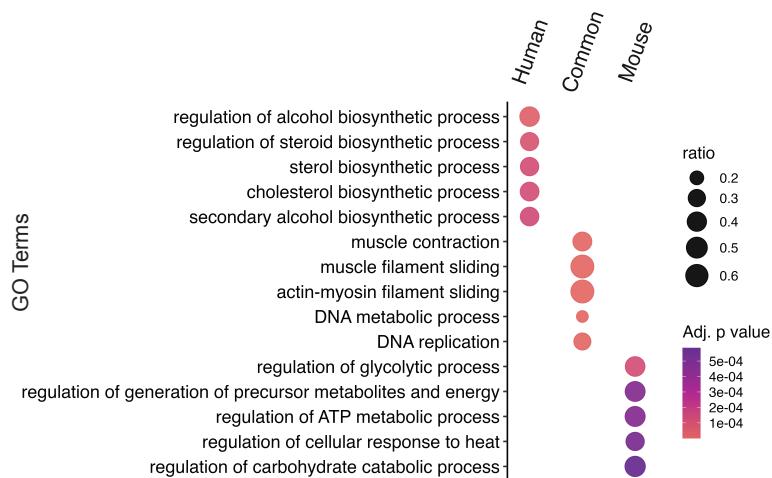
B



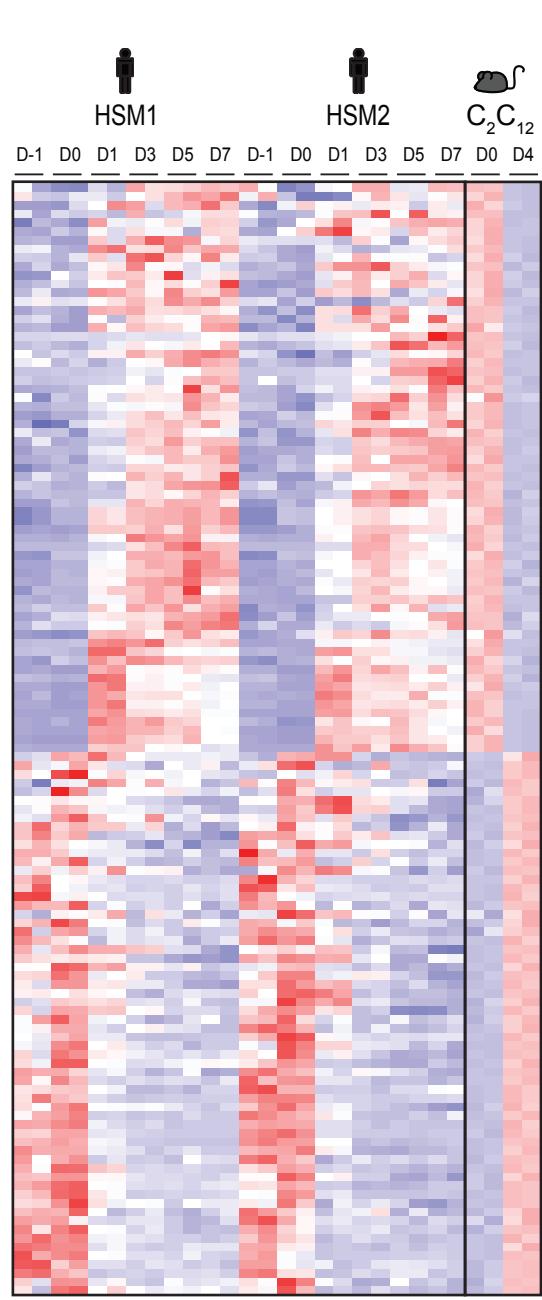
C



D



F



E

