# Epigenetic memory of Lateral Ganglionic Eminence derived hiPSCs cannot overcome barriers in authentic MSN differentiation

Bartley O. J. M.[1], Vinh N. N.[2], Lelos M.J.[1], Williams N. M.[2], Precious S. V.[1], & Rosser A. E.[1,3,4]

[1] *Brain Repair Group, School of Biosciences, Life Sciences Building, Cardiff University, Cardiff, United Kingdom*

[2] *Centre for Neuropsychiatric Genetics & Genomics, Division of Psychological Medicine & Clinical Neurosciences, Cardiff University, Cardiff, United Kingdom*

[3] *Neuroscience and Mental Health Research Institute, Hadyn Ellis Building, Cardiff University, Cardiff, United Kingdom*

[4] *Brain Repair And Intracranial Neurotherapeutics (B.R.A.I.N.) Biomedical Research Unit, College of Biomedical and Life Sciences, Cardiff University, Cardiff, United Kingdom*

## Abstract

Induced pluripotent stem cells (iPSCs) have previously been shown to retain some of the epigenetic features associated with the tissues from which they were derived, which in turn can help the iPSCs to differentiate towards similar cell fates. Here we investigate whether human iPSCs (hiPSC) derived from the lateral ganglionic eminence (LGE) retain sufficient epigenetic features of their original tissue to enhance their capacity to differentiate into striatal medium spiny neurons (MSN) compared to isogenic controls. We identify a subtle methylation signature within these isogenic hiPSCs, and observe line specific capacities to produce MSN-like cells and their subtypes. We directly compare these cells to authentic LGE derived MSNs and identify large differences in DNA methylation between these populations. Using single cell RNA sequencing, we also find subpopulations that do not reflect the transcriptional profile of authentic MSNs. While we observe some potential epigenetic "benefits" within our LGE derived hiPSC-MSNs (e.g. reduced hypomethylation of off-target pathways), we did not observe an improved propensity to differentiate towards an MSN-like fate. Overall, this work highlights that pluripotent stem cells are unique in their individual capacity for differentiation, that subtle differences between cell lines can have far

reaching effects on the final cell product, and that while hPSC-MSNs can model many aspects of MSN development, there are still epigenetic and transcriptional differences that limit the full recapitulation of the authentic LGE-derived MSN phenotype.

## Key words

Induced pluripotent stem cell; epigenetic memory; medium spiny neuron; lateral ganglionic eminence; stritatum; DNA methylation; single cell RNA seq; authenticity

## Introduction

Epigenetic mechanisms are critical mediators of gene expression and therefore responsible for the regulation of cellular fate. However, their precise roles throughout development are understudied and therefore underutilised. One important epigenetic mechanism is DNA methylation, which has been shown to undergo large shifts during development and is known to play a pivotal role in cellular phenotype specification through the regulation of gene expression (Greenberg, & Bourc'his, 2019; Smith & Meissner, 2013; Moore, Le, & Fan, 2013). Induced pluripotent stem cells (iPSC) have previously been shown to retain some of the epigenetic features associated with the tissues from which they were derived, sometimes termed an "epigenetic memory" of their tissue of origin (Kim et al., 2010). This epigenetic memory influences gene expression within these iPSCs, replicating some of the signature gene expression patterns of the original tissue and thereby encouraging differentiation towards phenotypes related to their tissue of origin (Table S1). Considering this, it is possible that iPSCs derived from cells already undergoing normal developmental processes towards a specific fate may retain some of this developing epigenome and remain epigenetically primed to differentiate towards that cell fate, while still retaining the benefits of a pluripotent cell source (such as indefinite proliferation). Here, we explore this possibility, using striatal medium spiny neurons (MSN) as our target phenotype.

MSNs are the principal neuronal cell type of the striatum and are the cell type predominantly damaged in Huntington's disease (HD) (Bates, Tabrizi & Jones, 2014). Considerable effort has been made to optimise protocols for differentiating human PSCs (hPSC) to an MSN phenotype, but despite these efforts, reliable generation of accurately specified hPSC derived MSNs (hPSC-MSN) remains elusive. Indeed, published protocols are generally inefficient, unable to produce homogeneous MSN populations, and are challenging to replicate between labs (Le Cann et al., 2021, Bartley et al., 2021, Garcia Jareno et al.,

2022). Currently, the successful differentiation of hPSCs to a striatal MSN phenotype is primarily determined through identification of a small number of markers, such as the co-expression of BCL11B (CTIP2) and PPP1R1B (DARPP-32). However, these markers are not unique to MSNs and their co-expression does not definitively indicate that a cell has successfully recruited all the key characteristics of an authentic and functional MSN phenotype. This is most evident in cell transplantation studies, in which hPSC-MSNs that express these markers do not yet consistently produce functional recovery comparable to that achieved using sources of *authentic* MSNs such as those derived from the developing fetal striatum (Bartley et al., 2021, Garcia et al., 2022; Lelos et al., 2016). It is therefore desirable to find strategies to overcome these differences in order to improve the accuracy and reliability of hPSC differentiation protocols, and thereby improve the quality and functional capacity of hPSC-MSNs for regenerative medicine and *in vitro* modelling.

Previously, we established that hiPSCs derived from human fetal whole ganglionic eminence (hWGE, the origin of striatal precursors, including MSNs, in normal development) are capable of differentiating towards striatal MSN-like phenotypes, and that these cells share commonly sought MSN features with hWGE tissues both *in vitro* and *in vivo* (Choompoo et al., 2021). However, it remains unclear whether these cells “benefit” from an epigenetic memory of tissue of origin, and to what extent this may impact the final phenotype of these cells, particularly in relation to the fetal tissues from which they were derived. In the present study, we addressed this by asking whether isogenic hiPSCs that have been derived from different somatic tissues of origin have the same propensity to be differentiated into striatal MSNs, and how closely MSNs from these lines resembled “authentic” MSNs differentiating within human fetal LGE, or those differentiated from a human ESC line.

# Results

## Generation and validation of isogenic hiPSCs derived from fetal LGE, cortex, and fibroblasts

To ensure genomic variability between hiPSC lines did not impact on our study, we generated isogenic hiPSCs from a single fetal donor using a non-integrating sendai virus to deliver reprogramming factors. We used this method to generate hiPSCs from three discrete regions of the developing human fetus: (1) the lateral ganglionic eminence (LGE); a telencephalic structure that gives rise to striatal MSNs and thus is developmentally primed to produce cells committed to an MSN fate; (2) the developing fetal cortex, another telencephalic structure that produces neuronal cells but does not contribute to the striatum; and (3) dermal fibroblasts which are non-neural and are therefore not epigenetically primed for a neuronal or MSN fate (Figure 1A). Colonies of hiPSCs were assessed for successful reprogramming, following which five clones were selected for further analysis; three derived from the LGE (L1, L2, L11), and one each from the cortex (C6) and dermal fibroblasts (F1). These clones exhibited typical pluripotent cell and colony morphology, expressed critical pluripotent genes OCT4, SOX2, TRA-1-60, NANOG, and SSEA4, and were capable of producing progeny of all three germ layers (Figure S1), thus, demonstrating the successful reprogramming of these fetal tissues into hiPSCs.

## Isogenic hiPSC lines derived from different tissues exhibit variability in striatal gene expression, despite all achieving MSN characterisation according to classic cellular markers.

Following exposure to a standardised MSN differentiation protocol (Figure 1A, MSN differentiation; adapted from Arber et al. 2015, RepairHD), all isogenic hiPSC lines consistently produced neuron-rich populations (as identified by βIII-tubulin/MAP2 expression, Figure 1B, Figure S2A, S3A), which coexpressed classic MSN markers CTIP2 and DARPP32 (Figure 1B, S2B-C), indicating that these hiPSCs are all capable of producing MSN-like phenotypes (hiPSC-MSN).

However, subtle differences could be observed between hiPSC cultures, for example, hiPSC line C6 consistently yielded the purest βIII-tubulin expressing neuronal populations, and significantly more than hiPSC line L11 ($F_{1,10}$=7.16, *p=0.005*; C6: 96.5%, L11: 78.9%; Figure 1B, Figure S2), suggesting the cortical derived C6 line may have a greater propensity to produce neuronal fates under these differentiation conditions. Simultaneously, we observed a reduction in DLX2, and an increase in NKX2.1 in C6 compared to other lines during the progenitor phase of our differentiation (DIV 20, Figure 1C). In striatal development DLX2 is expressed throughout the ganglionic eminence but critically has been shown to play a pivotal role in the formation of GABAergic neurons including MSNs (Petryniak et al., 2007; Victor et al., 2014), whereas NKX2.1 is preferentially expressed in the medial ganglionic eminence (MGE), and is common to the inteneuron progenitors migrating to the cortex (Wonders and Anderson, 2006).

Differences were also observed between fibroblast derived F1 cultures and the other hiPSC lines. ZNF503 (NOLZ1), expressed by LGE progenitors migrating to the developing striatum (Urbán et al, 2010) was expressed at a significantly higher level by F1 derived hPSC-MSNs compared to those derived from L11 (but not L2) cultures ($F_{4,10}$=3.54, *p=0.048*; Figure 1C). Across all time points, F1 derived differentiations expressed higher levels of CTIP2 (expressed by ~95% of DARPP32 expressing striatal MSNs; Arlotta et al 2008) than other hiPSC derived cultures ($F_{4,10}$=12.15, *p=0.001*; Figure 1C). Following maturation to DIV 40 F1 cultures also expressed more FOXP1 (expressed by ~100% of DARPP32 expressing striatal MSNs, Precious et al., 2016; $F_{8,20}$=5.82, *p=0.001*; Figure 1C) and DARPP32 (expressed by the majority of striatal MSNs, Ouimet et al., 1998; $F_{8,20}$=11.48, *p<0.001*; Figure 1C). The majority of mature MSNs can be subdivided by their functional response to dopamine into D1, D2 or D1/D2 subtypes (Gerfen et al., 1990; Lester et al.,

1993). While we observed no line specific differences in the D1 receptor gene DRD1, F1 derived hPSC-MSNs consistently expressed significantly higher levels of the D2 receptor gene DRD2 than other lines at DIV 30 and 40 ($F_{8,20}$=7.17, $p<0.001$; Figure 1C) potentially indicating greater yield of D2 type MSNs.

Notably, few differences were found between hPSC-MSNs derived from the two LGE derived hiPSC lines examined here, which showed a similar capacity to express early, mid, and late MSN genes. However, CALB1 (Calbindin, a mature striatal marker of matrix MSNs, Gerfen et al., 1985) was expressed at higher levels in mature L2 derived cultures compared to L11 cultures ($F_{8,20}$=7.40, $p<0.001$; Figure 1C).

Overall, despite sharing a uniform genetic background, this data demonstrate that these isogenic hiPSCs, derived from different fetal tissues, exhibit variations in their capacity to produce hPSC-MSNs, and that these differences are consistent over multiple rounds of differentiation. This data also suggests that these differences may be specific to tissue of origin, as differences between LGE derived hiPSC cultures were less pronounced than those observed between the lines derived from other tissues.

## hiPSCs accrue DNA methylation across striatal genes during reprogramming, which diminishes during *in vitro* differentiation towards MSN-like fates

Having observed that isogenic hiPSCs exhibit line-specific variability when undergoing an MSN differentiation, we next sought to characterise the methylome of our MSN-like cells. We conducted bulk DNA methylation analysis on hiPSC-MSN samples derived from tissues most likely to be most epigenetically distinct. Specifically, we reasoned that hiPSCs derived from LGE (L1, L2, and L11) would be "primed" for differentiation to an MSN phenotype, but that fibroblast derived hiPSCs (F1) would not be primed for differentiation to any central nervous system phenotype and thus would be the most epigenetically distinct of our hiPSCs. These were compared to undifferentiated controls, the primary fetal tissues from which these hiPSCs were derived, additional genomically distinct primary human LGE samples, and stage-equivalent cultures from a genetically distinct hESC line (H9) (Table S2; Figure S3A).

First we characterised the methylome of our undifferentiated hiPSCs. Hierarchical clustering revealed a separation of primary tissue from the derivative pluripotent hiPSC lines, indicating that the methylomes of undifferentiated hiPSCs were more similar to undifferentiated hESCs than to their somatic tissue of origin (Figure S3B). This difference was mostly driven by an increase in genomic methylation following reprogramming, but some regions that were previously hypermethylated in fetal tissues became hypomethylated during reprogramming (Figure S3C).

To better understand this methylome shift, we directly compared primary LGE tissues to our pluripotent L-hiPSCs. We identified 60197 significantly differentially methylated probes (DMP) between these cell types, which were associated with 3287 genes. Of these genes, 3088 had CpG sites which became more methylated following induction to pluripotency, conversely, only 308 had CpG sites which became less methylated (109 genes exhibited both an increase and decrease in methylation across associated CpG sites, Figure 2A, Table S3). Gene enrichment analysis revealed many of the genes which accrued methylation following reprogramming were associated with the striatum, neuronal processes and development, whereas genes which became demethylated were associated with pluripotent functions (Figure 2A). Overall, this demonstrates a removal of the striatal-specific

methylome, and restoration of a more naïve pluripotent state during induction to pluripotency.

Next, we examined how the methylome changed following *in vitro* differentation from a pluripotent phenotype to an MSN-like one, by comparing undifferentiated hPSCs to hPSC-MSNs. Here we found 81627 DMPs, indicating another widespread change across the methylome following differentiation. These DMPs were associated with 1064 genes, 717 of which accrued methylation during differentiation and 384 became less methylated (37 genes exhibited both an increase and decrease in methylation across associated CpG sites, Table S4). As above, genes that were demethylated in pluripotent cells were strongly associated with pluripotency phenotypes, cycles and functions. Whereas, genes which were more demethylated in hPSC-MSNs were associated with striatal tissues, and biological processes related to neural development and function, including canonical MSN genes such as MEIS2, FOXP2, and TLE4 (Figure 2B). Thus indicating that our MSN differentiation protocol specifically induces epigenetic changes at relevant genomic regions for the MSN phenotype.

## A stable tissue-specific methylation signature persists through reprogramming to pluripotency and through subsequent MSN differentiation

We next sought to establish whether an epigenetic signature of tissue of origin was still identifiable in the methylome of our hiPSCs following reprogramming, and understand how this influenced the final methylome of derirative MSNs. During earlier characterisation it became apparent that our hiPSCs were more similar to each other than to hESCs, likely reflecting their common genomic heritage and form of pluripotency (i.e. iPSC; Figure S2B-C). However, we also observed that hiPSCs derived from LGE tissues were more similar to each other than to those derived from the isogenic fibroblasts, indicating a common residual methylation signature likely resulting from tissue of origin (Figure 2C; Figure S2D). This methylation pattern also persisted in the hiPSC methylomes following differentiation to MSN-like cells (Figure 2D; Figure S2D), indicating that this methylation signature is stable within these lines and has become a line-specific signature.

Next, we explored the DMPs between L- and F- hiPSC-MSNs to better define this signature and explore how this could influence gene expression in MSN populations. Few DMPs were observed between the LGE-derived hiPSC-MSN lines, likely due to their shared tissue of origin (Figure S3E). Subsequently, for further comparisons, LGE-derived hiPSC-MSNs were pooled into one group (L-hiPSC-MSN). Between L-hiPSC-MSNs and F-hiPSC-MSNs there were 437 significant DMPs, which were associated with 16 genes (Figure 2E; Table S5). Of the 11 genes which exhibited reduced methylation in L-hiPSC-MSNs compared to F1-hiPSC-MSNs, five were associated with brain tissue; CNTNAP1, FAM135B, KCNK9, GABRG2 and CBLN4, and two of these (GABRG2 and CBLN4) are specifically associated with GABAergic neurons. Conversely, of the 7 genes which exhibited increased methylation in L-hiPSC-MSNs compared to F-hiPSC-MSNs, four (OR5AS1, OR8U8, OR10AG1, and MYORG) were associated with olfactory pathways and function. Furthermore, three of the genes observed to be more methylated in F-hiPSC-MSNs compared to L-hiPSC-MSNs are down regulated in the olfactory bulb (CNTNAP1, FAM135B and SMARCA2). Notably the dorsal LGE gives rise to non-striatal cells which contribute to the olfactory bulb (Stenman et al., 2003), and therefore expression of these genes may prove to be undesirable in hPSC-MSN populations. Overall, while the line specific differences between our hiPSC lines only affected a small number of genes, these differences do appear

to be relevant to the MSN phenotype and this suggests LGE derived hiPSCs may be marginally more suited to recreating the MSN epigenome.

## hPSC-MSNs do not reproduce the "authentic" methylome of primary LGE tissue regardless of tissue of origin

We next sought to confirm whether the subtle epigenetic signature observed in our LGE derived hiPSCs produced MSNs more similar to authentic phenotypes when compared to those achieved by either the fibroblast derived hiPSCs or epigenetically naïve hESCs. To achieve this, we compared the methylome of our hPSC-MSNs to authentic LGE-derived MSN phenotypes.

We identified over 100,000 DMPs between primary LGE and respective hPSC-MSNs, indicating large discrepancies in the methylome of these cell populations (Tables S6-9). Examination of the associated genes revealed 2231 genes were differentially methylated between primary LGE and hPSC-MSNs, with the vast majority (2224) exhibiting reduced methylation in primary LGE tissues compared to hPSC-MSNs (Figure 2F). These genes were strongly associated with brain tissues and striatal substructures (Figure 2G), many of which were also associated with neuronal development and activity (Figure S4).

Notably, LGE tissues comprise a heterogeneous population of cell types spanning different phenotypes and maturities, and therefore these differences may be partially driven by the variability of cell types in our primary LGE tissue group that were not present in the more focally fate-restricted hPSC-MSNs. Some enriched terms were associated with non-neural phenotypes (e.g. astrocytes, Figure S4). However, considering (i) that the majority of differentially methylated genes are associated with neural and striatal phenotypes and (ii) that we identified very few (7-34) genes uniquely demethylated in hPSC-MSNs, it is reasonable to assume that this epigenetic disparity was, at least in part, driven by a discordance between striatal and neural phenotypes in the primary LGE cultures compared to those within the hPSC-MSNs. Overall, this data suggests that the global increase in methylation observed following induction to pluripotency has not been fully reversed following directed differentiation to an MSN-like phenotype, and that while hPSC-MSNs do exhibit some methylation changes on striatal genes, this does not completely recapitulate an authentic striatal methylome.

## Single cell RNA sequencing reveals hPSC-MSNs are similar, but not equivalent, to authentic LGE-derived MSNs

To explore the biological impact of the epigenetic disparity observed above and further identify expression differences between isogenic hiPSC-MSNs, we performed scRNAseq of four samples: (i) primary human LGE (11 pcw, unrelated to hiPSCs); (ii) L2 hiPSC-MSNs (DIV 40); (iii) F1 hiPSC-MSNs (DIV 40); and (iv) H9 hESC-MSNs (DIV 40).

Within the primary LGE we identified transcriptional phenotypes spanning normal MSN development and we sub-clustered cells of the MSN lineage for characterisation, using hallmark genes to identify specific stages of MSN development and subtypes (Figure 3 A-D, Figure S5. Specifically, we observed a cluster of cells positive for genes including HES1, SOX2, and NES, typical of apical progenitors of the LGE ventricular zone (Figure 3 D). The neighbouring cluster was positive for genes such as TOX3, ASCL1, and DLX2 which are each associated with basal progenitors of the subventricular zone of the LGE (Figure 3 D). The remaining clusters were all strongly positive for global MSN genes including MEIS2 and BCL11B, and were GABAergic as demonstrated by positive expression of GAD1 and GAD2 (Figure 3 C-D). The closest cluster to the basal progenitors was uniquely positive for POU3F1, which has previously been identified as a stage specific marker for pre-MSNs undergoing subtype fate decisions (Bocchi et al., 2021). However, POU3F1 was only expressed by a small proportion of this cluster, instead other genes such as LMO3 better defined this cluster in our analysis (Figure 3 C-D). Further assessment of this population through additional sub-clustering revealed a division into three populations which could be defined by early subtype MSN markers into (i) uncommitted progenitors, (ii) D1/EBF1+ progenitors, and (iii) D2/SIX3+ progenitors (Figure S5). Additionally, pseudotime analysis (Figure 3 B) revealed a fate decision occurring within this population which later leads to MSN subtype specification, further supporting the identification of this cluster as pre-MSN progenitors. Two further clusters expressed D1 subtype markers including ISL1 and EBF1 (Figure 3 C-D), suggesting they were of the D1 MSN lineage. Across these cells we observed a gradient increase in genes associated with mature D1-MSNs such as GRIK2, TAC1 and IKZF1, indicating these as immature and mature D1 MSNs (Figure 3 D). Finally, we observed three additional clusters which branched off from the D2/SIX3+ progenitors of the pre-MSNs. These clusters contained cells positive for early D2 progenitors such as SIX3 and SP9 (Figure 3 C), but could be differentiated by several genes, including CLMP, CXCL12,

and ADARB2 (Figure 3 D). The ADARB2+ cluster also expressed TSHZ1 and ERBB4, suggestive of a D1/D2 MSN phenotype (also called eccentric MSNs). Finally, we observed D1, D2, and D1/D2 MSNs could be identified by their unique expression of FOXP1 and FOXP2, D1 MSNs co-expressing both genes, D2 MSNs expressing just FOXP1, and D1/D2 MSNs expressing just FOXP2 (Figure 3 C-D).

Similar analysis was conducted on each population of hPSC-MSNs (Figure 3 E-G, Figure S6-8). We observed an enriched MSN phenotype in our hPSC-MSNs, exhibiting similar patterns of expression as in the primary LGE MSNs. For instance, global MSN markers BCL11B and MEIS2 were enriched across the populations, as were GABAergic marker genes GAD1 and GAD2, indicating a strong commitment to an MSN-like phenotype. Markers consistent with early progeny and pre-MSNs (e.g. HES1, TOX3, NES, ASCL1, DLX2 and LMO3) were present within the populations at this late time point, but only in a low proportion of cells, indicating the majority had committed to a more mature phenotype. Across all three hPSC-MSN populations we observed enrichment of MSN subtype genes (e.g. EBF1, TAC1, SIX3, SP9), and markers of maturing phenotypes (e.g. GRIK2, RELN, IKZF1, GRIK3, CLMP, CXCL12, ADARB2), indicating hPSC-MSNs had successfully differentiated to the similar subtypes as authentic LGE derived MSNs. However, D1 subtypes (BCL11B/MESI2/EBF1+ cells) were underrepresented in our differentiations, and conversely D1/D2 phenotypes (BCL11B/MEIS2/ADARB2+ MSNs) made up a larger proportion of the final cell population when compared to primary LGE derived MSNs, particularly in the L-iPSC-MSNs. Notably, subgroup distinctions were less clear compared to primary tissues, producing less defined clusters. This may be due to the reduction in representation of different developmental stages, as the hPSC cultures were collected at the same time point (DIV40) making them more homogeneous, wheras the fetal LGE tissue contains a greater range of cell stages from the surface of the ventricular zone to the mantle zone and developing striatum. Yet, some expression patterns and borders were still indistinct regardless of this limitation, for example, some genes which were restricted to the D1/D2 subtype in fetal derived MSNs (e.g. ADARB2) were also observed across the D2-like MSNs in all three hPSC-MSN samples. However, pseudotime analysis identified similar trajectories between cell lines that loosely followed the developmental path from LMO3 expressing progenitors to the mature phenotypes, with D1, D2 and D1/D2 phenotypes branching off into separate phenotypes, indicating these as discrete cell populations (Figure 3 E-G).

Next we created an integrated dataset comprised of all four samples, including non-MSN phenotypes collected from the fetal LGE tissue, for further comparisons (Figure 4). We started by comparing the transcriptional differences between the isogenic L2- and F1- hiPSC-MSN populations, to explore whether their unique epigenomes altered their final phenotypes. Between these two populations, there were 89 genes expressed at significantly different levels, 31 higher in the F1-hiPSC-MSNs and 58 higher in the L2-hiPSC-MSNs. These differences primarily reflected a difference in subtype proportions, for example, greater levels of FOXP1, ZNF503, SIX3, CLMP, and GRIK3 expression were observed across the F1-hiPSC-MSNs, indicating higher yields of both D1-like and D2-like MSN subtypes than in the L2-hiPSC-MSNs; conversely, greater proportions of ADARB2, TSHZ1, FOXP2, ERBB4, and LUZP2 expressing cells were observed in L2-hiPSC-MSNs, indicating a higher yield of D1/D2-like MSNs than in the F1-hiPSC-MSNs (Figure 4 E-G). Further exploration of specific MSN subtypes did not reveal gene expression profiles unique to individual lines, indicating that the epigenomic background of our hiPSC lines mainly effected their potential to yield total MSNs and subtype fate commitment. Specifically this caused the F1 line to yield greater proportions of D1- and D2-like MSNs, and the L2 line to yield greater numbers of ADARB2+ D1/D2-like MSNs, in spite of their isogenic nature.

Lastly, we used this opportunity to explore how these three hPSC-MSN populations compared to authentic LGE derived MSNs. Overall, there was considerable overlap between hPSC-MSNs and fetal LGE, suggesting they are capable of producing similar phenotypes (Figure 4 A-D). Yet, there were some notable exceptions, specifically we observed six clusters (2, 3, 4, 10, 11, & 16) to be primarily comprised of hPSC-MSNs (Figure 4D). Clusters 2 and 3 expressed high levels of CLMP, potentially indicating these are CLMP expressing D2 MSNs. However only cluster 3 expressed established D2 genes SIX3 and SP9, indicating that cluster 2 may possess an incomplete or abberant phenotype. Comparisons of cluster 3 (CLMP/SIX3/SP9+ but containing ~0% Primary LGE cells) to cluster 15 (which also contained CLMP/SIX3/SP9+ cells and contained ~59% primary LGE cells) revealed 292 differentially expressed genes: cells in cluster 3 expressed genes associated with neuronal projections and axon regulation, whereas cells in cluster 15 expressed genes associated with fetal brain tissues, indicating that this division is likely due to differences in maturity between these cells. Similarly, cells in cluster 10 expressed D1 marker EBF1, but did not express genes that differentiated between immature and mature D1 cells in the primary LGE tissues (e.g. ISL1 or RELN), potentially indicating that these cells are an incomplete D1 phenotype.

Whereas clusters 4, 11, and 16 contained cells positive for D1/D2 genes such as ADARB2, which only represented a small proportion of the overall population in primary LGE, likely emphasising a bias for our protocol to over-produce the D1/D2 phenotype (Figure 4 F).

# Discussion

hPSCs are a powerful research tool due to their ease of accessibility and potential to differentiate into any cell phenotype in the body. However, it has become increasingly clear that variability exists between lines and types of hPSCs (Ortmann and Vallier, 2017), and understanding the depth, range and causes of these differences is important for their effective use. Here, we sought to explore how epigenetic variability between isogenic hiPSCs can influence their differentiation capacity, and specifically whether hiPSCs derived from LGE tissues would benefit from an epigenetic memory of tissue of origin compared to hiPSCs derived from other tissues i.e. by increasing their propensity to differentiate towards an MSN-like phenotype.

We have identified subtle but clear epigenetic signatures in our hiPSCs and shown that these signatures persist through differentiation into hPSC-MSNs. Our data, combined with other published reports exploring epigenetic memory in neural derived iPSCs (Tian et al., 2011, Hargus et al., 2014, Roost et al., 2017), suggests that this epigenetic signature is retained from the original tissue of origin. We have further evidenced that these isogenic hiPSCs vary in their capacity to produce MSN-like cells when undergoing a standardised differentiation; that these differences are consistent over multiple differentiations within a given line; and that their final MSN products exhibit different gene expression profiles and different proportions of MSN subtypes. As these hiPSCs share a genetic heritage, we can confidently attribute this variable differentiation capacity to their line specific epigenetic variability, which includes the epigenetic signature inherited from their tissue of origin.

Other works exploring the residual epigenome following iPSC induction typically observe an innate bias to return to cellular phenotypes related to the iPSCs tissue of origin. For example, Kim et al. 2010 demonstrated iPSCs derived from mouse bone marrow more efficiently formed haematopoietic colonies than iPSCs derived from mouse fibroblast cells, and that the converse was true for the efficient formation of osteogenic colonies. This has also been reported in studies exploring iPSCs derived from neural tissues: Tien et al. 2011 reported iPSCs derived from mouse astrocytes yielded more TH expressing neurons than mouse fibroblast derived iPSCs; Hargus et al. 2014 reported hiPSCs derived from fetal neural stem cells produced neural progenitors that were more enriched for neural genes and showed improved survival following transplant into the mouse brain compared to hiPSCs derived from cord blood and fibroblasts; and Roost et al. 2017 showed fetal brain derived hiPSCs

retained brain-specific DNA methylation and had an improved propensity to differentiate into neural cells. While we did observe some potential epigenetic "benefits" within our LGE derived hiPSC-MSNs (e.g. reduced hypomethylation of off-target pathways), and arguably an enhanced capacity for neural commitment in the C6-hiPSCs (consistently yielding the highest proportion of neuronal cells), we did not observe an improved propensity to differentiate towards a D1 or D2 MSN-like fate. In fact, across multiple differentiations we observed that mature LGE derived hiPSC-MSNs expressed global MSN genes (e.g. CTIP2 and DARPP32) at lower levels than fibroblast derived hiPSC-MSNs produced in parallel, and potentially a bias to produce the less studied D1/D2 eccentric MSN subtype instead. Notably, the differentiation protocol used in this study was primarily developed and optimised for naïve hESCs (Arber et al., 2015). It is therefore possible that the comparative success of our fibroblast derived hiPSCs over our other hiPSC lines was because the dosages and exposure times used for pathway signalling modulators are better suited to cell lines naïve to neuronal and striatal development. If so, this would have created sub-optimal conditions for our "primed" LGE derived hiPSCs causing them to either yield fewer MSNs, or to bias them towards the D1/D2 subtype. Understanding exactly how and why these lines responded differently to the same protocol would require substantial further investigation.

DNA methylation is a major epigenetic modifier that plays a central role in the regulation of gene expression, capable of both silencing and promoting gene expression depending on the locus of methylation. During normal development, DNA is methylated and demethylated in a dynamic process that guides cells to their final phenotype. A fully differentiated cell exhibits a stable DNA methylation pattern that regulates tissue-specific gene transcription (Moore, Le, & Fan, 2013). Adult cells that are returned to a pluripotent state (as iPSCs) exhibit large scale DNA methylation events, causing the majority of the genome to return to a methylated state and become comparable to undifferentiated ESCs (Papp and Plath, 2013). Here, we observed this substantial increase in DNA methylation in LGE derived hiPSCs compared to fetal LGE tissues, and many of the genes which acquired methylation were associated with striatal tissues indicating a general loss of LGE tissue-specific methylome. By differentiating these hPSCs into hPSC-MSNs, we observed further dynamic shifts in DNA methylation across several genes associated with striatal tissues, suggesting the recapitulation of normal striatal methylation and thus striatal differentiation. However, the methylome of hPSC-MSNs remained substantially different from authentic fetal LGE tissues, and many of the genes that remained methylated in hPSC-MSNs were

associated with striatal tissues and neural function, showing that hPSC-MSNs have recruited part of, but probably not all of, a striatal methylome. Overall, this indicates (i) that this and similar differentiation protocols are insufficient to fully recreate an authentic MSN-like methylome, and (ii) that LGE derived hiPSCs do not retain enough of their original methylome to drastically alter their capacity for later MSN differentiation.

Interpreting the importance of the stark differences in these methylomes is challenging as (i) the precise role of DNA methylation is not fully understood across every gene and associated genomic region (Greenberg, & Bourc'his, 2019; Smith & Meissner, 2013; Moore, Le, & Fan, 2013), and (ii) the fetal LGE tissues studied here are heterogeneous, having been shown to contain cells at various developmental stages committing to different cell fates (Bocchi et al., 2021; Shi et al., 2021). A fuller understanding could be achieved if the methylome of these two populations was explored at the single cell level. As such technology becomes more widely available, we anticipate future comparisons will further elucidate the reason for the observed discordance between authentic and hPSC derived MSNs. Yet, our scRNAseq analysis allowed us to remove much of the heterogeneity of primary LGE tissues by isolating just those cells relevant to the MSN lineage. Here we observed that, despite their incomplete methylome, hPSC-MSNs and their derivative subtypes exhibit transcriptional signatures comparable to authentic fetal derived MSNs and that at least some of the differences observed between hPSC-MSNs can be attributed to proportional differences in MSN subtype yield. However, we also identified some clusters of hPSC-MSNs that did not produce typical expression patterns of authentic MSNs, potentially indicating these cells have not achieved an authentic phenotype. Overall, this work highlights that (i) hPSC lines are unique in their individual capacity for differentiation, (ii) that subtle differences that exist between lines can have far reaching effects on the final cell product and (iii) that hPSC products which express canonical genes may still include some incomplete or inauthentic cells within their population, empahsising the need to rigorously validate hPSC models.

## Conflicts of Interest

The authors declare no conflicts of interest.

## Funding declarations and acknowledgements

This work was funded by Wellcome Trust studentship grants 105215/Z/14/Z and 105215/Z/14/A, MRC grant 518008, and a Cardiff University NMHRI seedcorn grant. The authors would like to acknowledge and thank the technical expertise of the SWIFT team in the Brain Repair Group at Cardiff University and at University Hospital Wales, and the Genomics Research HUB at Cardiff University, Lesley Bates and Joanne Morgan.

## Author Contributions

Conception and design of the study: OJMB, NNV, NMW, SVP and AER. Acquisition of data: OJMB. Analysis and interpretation of data: OJMB. Drafting or revising the manuscript: OJMB, MJL, SVP, and AER. All authors have approved the final article.

## Methods

### Primary Tissue Culture

All primary human fetal tissue was acquired from the South Wales Initiative for Fetal Transplantation (SWIFT) research tissue bank. It was dissected and dissociated as previously described (Roberton et al., 2018).

### Generation of hiPSCs

Primary tissues were cultured on PDL (Gibco) treated nunc cell culture plates (Thermofisher). Neural tissues were maintained in a neural media (DMEM/F12, +2% B27 Supplement 50x (B27; Gibco), +1% fetal bovine serum (FBS; Gibco), +1% Penicillin-Streptomycin (PS; Gibco)), and fibroblasts were cultured in a basic media (DMEM/F12, +10% FBS, +1% MEM Non-Essential Amino Acids Solution 100x (MEM-NEAA, Gibco), +1% L-Glutamine (Gibco), +1% PS). 3 days following dissociation and plate down we induced pluripotency via the CytoTune -iPS 2.0 Sendai Reprogramming Kit (Invitrogen), at MOI = 5 for 24 hours. Tissues were maintained under pluripotent culture conditions from day 5 post-infection.

### Culture of hPSCs

hPSCs (hiPSCs and hESC line H9 (WA09, Wicell)) were maintained on Geltrex (Gibco) or hESC-qualified Matrigel (Corning) coated nunc plates, in E8 or E8-Flex media (Gibco).

hPSCs were passaged when between 70-90% confluency using EDTA (Sigma) or ReLeSR (StemCell Technologies).

## Spontaneous differentiation of hPSCs

hiPSCs at 40% confluency were cultured in KSR medium (KnockOut DMEM (Gibco), +10% KnockOut Serum Replacement (Gibco), +1% L-glutamine, +1% MEM-NEAA, +1% PS, +0.1% β-mercaptoethanol (Gibco)) for 5 days to allow for spontaneous differentiation.

## Directed differentiation of hPSCs towards MSN-like fate

hPSCs were differentiated towards an MSN-like fate using a protocol adapted from previous works (Arber et al., 2015, Repair HD; Figure 1A).

Prior to differentiation, hPSCs were cultured on Growth Factor Reduced Matrigel (Corning; diluted 1:15 in DMEM-F12 and coated for 1 hour at 37°C) in E8 or E8-Flex media until ~70% confluency (day 0). Differentiation was initiated by changing media to N2B27-RA media (2:1 ratio of DMEM/F12 and Neurobasal medium (Gibco), +1% L-glutamine, +0.6% N2 supplement 100x (N2; Gibco), +0.6% B27 without retinoic acid supplement 50x, (B27-RA; Gibco), +0.2% MycoZap plus CL (Lonza), +0.1% β-mercaptoethanol), supplemented with 10 μM SB-431542 (SB; Tocris), 200 nM dorsomorphin (R&D),100 nM LDN-193189 (LDN; Tocris) and 2 μM XAV-939 (Xav; Tocris). From day 6, SB was removed from media. On day 7, cultures were passaged using EDTA at a 2:3 ratio split onto PDL + Fibronectin (Millipore) coated nunc plates. From day 10, dorsomorphin and LDN were removed from media, and 25 ng/ml Activin A (R&D) was added. On day 16, cultures were passaged using Accutase (Sigma) and frozen into N2B27-RA media supplemented with 10% DMSO (Sigma) and stored at -80$^{o}$C. Following thawing, cells were plated at ~100k cells/cm$^2$ on PDL+Laminin (Sigma) coated nunc plates in N2B27-RA media supplemented with Xav and Activin. From day 18, Xav and Activin were removed from media. From day 20, media was replaced with N2B27 (as N2B27-RA, but B27-RA is replaced with B27), supplemented with 10 ng/ml BDNF (Peprotech) and 10ng/ml GDNF (Peprotech). Medium was replaced every second (N2B27-RA) or third (N2B27) day throughout. Cells were maintained up to day 40.

## Fluorescent immunocytochemical staining and analysis

Cells were fixed in 3.75% PFA (Millipore) and permeabilised in 100% molecular grade ethanol (Sigma). Fluorescent immunocytochemistry was performed using standard protocols, including incubations in blocking solution, primary antibodies, and secondary antibodies. Primary antibodies: α-FP (Abcam), α-SMA (Sigma), βIII-tubulin (Sigma), CTIP2 (Abcam), DARPP32 (Abcam),DESMIN (Abcam), MAPa/b (Sigma), NANOG (Abcam), NESTIN (Neuromics), OCT4 (Abcam), SOX2, (Abcam), SSEA3/4 (Abcam), TRA-1-60 (Abcam), Vimentin (Millipore).

Stained cells were visualised using either a ZEISS Axio Imager 2 upright fluorescent microscope with Axiovision software, or using a Leica DMI6000B inverted fluorescent microscope with LAS X software. Cell counts were conducted using 5 images per sample and three samples per condition. Statistical analysis was conducted using SPSS (IBM).

## Gene expression analysis using PCR and qPCR

RNA was extracted using RNeasy Mini-Prep (Qiagen). cDNA was synthesised using SuperScript IV First-Strand Synthesis System (Thermofisher). PCR was conducted using BioTaq DNA polymerase (Bioline). qPCR was conducted using PowerUp SYBR Green (Invitrogen). For qPCR, data is expressed as ΔΔCt, normalised to the average Ct of two reference genes (GAPDH and β-ACTIN), and to the gene expression in H9-MSNs at day 30. Fold change expression values were used for statistical analysis. Statistical analysis was conducted using SPSS (IBM).

| Gene | Forward Primer (5'->3') | Reverse Primer (5'->3') |
|---|---|---|
| β-ACTIN | GCTCTTTTCCAGCCTTCCTT | CGGATGTCAACGTCACACTT |
| GAPDH | GTCGGAGTCAACGGATTTGG | ATGGAATTTGCCATGGGTGG |
| OCT4 | GTGGAGGAAGCTGACAACAA | ATTCTCAGGTTGCCTCTCA |
| SOX2 | CAAGATGCACAACTCGGAGA | GCTTAGCCTCGTCGATGAAC |
| DLX2 | AGCAGCTATGACCTGGGCTA | TCCTTCTCAGGCTCGTTGTT |
| NKX2.1 | ATGTCGATGAGTCCAAAGCA | CTCCATGCCCACTTTCTTGT |
| NOLZ1 | CTGCAGCCCCTGCCTTCCAC | GTCGGGCTTCCCGATCTGCG |
| CTIP2 | CCATCCTCGAAGAAGACGAG | ATTTGACACTGGCCACAGGT |
| FOXP1 | CGATCCCTTCTCTGATTTGC | CATGCATAATGCCACAGGAC |
| FOXP2 | CCACGAAGACCTCAATGGTT | GCTCTTCCTTGACGTGGATT |
| DARPP32 | CTGGGCAAAAGACAACCTGT | GGTCTTCCACTTGGTCCTCA |
| CALB1 | GATACTGACCACAGTGGCTTC | TGCACTGGTAGTAACCTGGC |

| | | |
|---|---|---|
| DRD1 | AGCGAAGTCCACATTCCAAG | ATGTCTTCTCGCTCCTCCAA |
| DRD2 | TCCTGAACTTGTGTGCCATC | GATGGAGATCATGACGGTGA |

## DNA Methylation Analysis

DNA methylation analysis was conducted as previously described (Choompoo et al., 2021).

## scRNAseq Analysis

scRNAseq performed using the Chromium Single Cell 30 v2 platform (10X Genomics). For each sample ~10,000 viable cells were collected for library generation. Libraries were sequenced with the Illumina NovaSeq platform, using paired end sequencing with single indexing. Raw data files (Fastq) were demultiplexed, processed, aligned and quantified using the Cell Ranger Single-Cell Software Suite (10x Genomics).

Downstream analysis was performed using Seurat5. Cells with <200 or >7500 genes were removed, as were those with >10% mitochondrial gene expression. For analysis of primary LGE alone, cell clusters expressing non-MSN lineage markers were removed. Pseudotime analysis was conducted using Monocle3.

## Gene enrichment analysis

Gene sets which were found to be differentially methylated (DNA methylation) or differentially expressed (scRNAseq) were analysed for significant terms using Enrichr (Chen et al., 2013, Kuleshov et al., 2016).

# References

Greenberg, M. V., & Bourc'his, D. (2019). The diverse roles of DNA methylation in mammalian development and disease. *Nature reviews Molecular cell biology*, *20*(10), 590-607.

Smith, Z. D., & Meissner, A. (2013). DNA methylation: roles in mammalian development. *Nature Reviews Genetics*, *14*(3), 204-220.

Moore, L. D., Le, T., & Fan, G. (2013). DNA methylation and its basic function. *Neuropsychopharmacology*, *38*(1), 23-38.

Kim, K., Doi, A., Wen, B., Ng, K., Zhao, R., Cahan, P., ... & Daley, G. Q. (2010). Epigenetic memory in induced pluripotent stem cells. *Nature*, *467*(7313), 285-290.

Bates, G., Tabrizi, S., & Jones, L. (Eds.). (2014). *Huntington's disease* (No. 64). Oxford University Press.

Le Cann, K., Foerster, A., Rösseler, C., Erickson, A., Hautvast, P., Giesselmann, S., ... & Lampert, A. (2021). The difficulty to model Huntington's disease in vitro using striatal medium spiny neurons differentiated from human induced pluripotent stem cells. *Scientific reports*, *11*(1), 6934.

Bartley, O. J., Lelos, M. J., Gray, W. P., & Rosser, A. E. (2021). Do foetal transplant studies continue to be justified in Huntington's disease?. *Neuronal Signaling*, *5*(4), NS20210019.

Jareño, P. G., Bartley, O. J., Precious, S. V., Rosser, A. E., & Lelos, M. J. (2022). Challenges in progressing cell therapies to the clinic for Huntington's disease: A review of the progress made with pluripotent stem cell derived medium spiny neurons. *International Review of Neurobiology*, *166*, 1-48.

Lelos, M. J., Roberton, V. H., Vinh, N. N., Harrison, C., Eriksen, P., Torres, E. M., ... & Dunnett, S. B. (2016). Direct comparison of rat-and human-derived ganglionic eminence tissue grafts on motor function. *Cell Transplantation*, *25*(4), 665-675.

Choompoo, N., Bartley, O. J., Precious, S. V., Vinh, N. N., Schnell, C., Garcia, A., ... & Rosser, A. E. (2021). Induced pluripotent stem cells derived from the developing striatum as a potential donor source for cell replacement therapy for Huntington disease. *Cytotherapy*, *23*(2), 111-118.

Arber, C., Precious, S. V., Cambray, S., Risner-Janiczek, J. R., Kelly, C., Noakes, Z., ... & Li, M. (2015). Activin A directs striatal projection neuron differentiation of human pluripotent stem cells. *Development*, *142*(7), 1375-1386.

Petryniak, M. A., Potter, G. B., Rowitch, D. H., & Rubenstein, J. L. (2007). Dlx1 and Dlx2 control neuronal versus oligodendroglial cell fate acquisition in the developing forebrain. *Neuron*, *55*(3), 417-433.

Victor, M. B., Richner, M., Hermanstyne, T. O., Ransdell, J. L., Sobieski, C., Deng, P. Y., ... & Yoo, A. S. (2014). Generation of human striatal neurons by microRNA-dependent direct conversion of fibroblasts. *Neuron*, *84*(2), 311-323.

Wonders, C. P., & Anderson, S. A. (2006). The origin and specification of cortical interneurons. *Nature Reviews Neuroscience*, *7*(9), 687-696.

Urbán, N., Martín-Ibáñez, R., Herranz, C., Esgleas, M., Crespo, E., Pardo, M., ... & Canals, J. M. (2010). Nolz1 promotes striatal neurogenesis through the regulation of retinoic acid signaling. *Neural development*, *5*, 1-22.

Arlotta, P., Molyneaux, B. J., Jabaudon, D., Yoshida, Y., & Macklis, J. D. (2008). Ctip2 controls the differentiation of medium spiny neurons and the establishment of the cellular architecture of the striatum. *Journal of Neuroscience*, *28*(3), 622-632.

Precious, S. V., Kelly, C. M., Reddington, A. E., Vinh, N. N., Stickland, R. C., Pekarik, V., ... & Rosser, A. E. (2016). FoxP1 marks medium spiny neurons from precursors to maturity and is required for their differentiation. *Experimental neurology*, *282*, 9-18.

Ouimet, C. C., Langley-Gullion, K. C., & Greengard, P. (1998). Quantitative immunocytochemistry of DARPP-32-expressing neurons in the rat caudatoputamen. *Brain research*, *808*(1), 8-12.

Gerfen, C. R., Engber, T. M., Mahan, L. C., Susel, Z. V. I., Chase, T. N., Monsma Jr, F. J., & Sibley, D. R. (1990). D1 and D2 dopamine receptor-regulated gene expression of striatonigral and striatopallidal neurons. *Science*, *250*(4986), 1429-1432.

Lester, J., Fink, S., Aronin, N., & DiFiglia, M. (1993). Colocalization of D1 and D2 dopamine receptor mRNAs in striatal neurons. *Brain research*, *621*(1), 106-110.

Gerfen, C. R., Baimbridge, K. G., & Miller, J. J. (1985). The neostriatal mosaic: compartmental distribution of calcium-binding protein and parvalbumin in the basal ganglia of the rat and monkey. *Proceedings of the National Academy of Sciences*, *82*(24), 8780-8784.

Stenman, J., Toresson, H., & Campbell, K. (2003). Identification of two distinct progenitor populations in the lateral ganglionic eminence: implications for striatal and olfactory bulb neurogenesis. *Journal of Neuroscience*, *23*(1), 167-174.

Bocchi, V. D., Conforti, P., Vezzoli, E., Besusso, D., Cappadona, C., Lischetti, T., ... & Cattaneo, E. (2021). The coding and long noncoding single-cell atlas of the developing human fetal striatum. *Science*, *372*(6542), eabf5759.

Ortmann, D., & Vallier, L. (2017). Variability of human pluripotent stem cell lines. *Current opinion in genetics & development*, *46*, 179-185.

Tian, C., Wang, Y., Sun, L., Ma, K., & Zheng, J. C. (2011). Reprogrammed mouse astrocytes retain a “memory” of tissue origin and possess more tendencies for neuronal differentiation than reprogrammed mouse embryonic fibroblasts. *Protein & cell*, *2*(2), 128-140.

Hargus, G., Ehrlich, M., Arauzo-Bravo, M. J., Hemmer, K., Hallmann, A. L., Reinhardt, P., ... & Zaehres, H. (2014). Origin-dependent neural cell identities in differentiated human iPSCs in vitro and after transplantation into the mouse brain. *Cell Reports*, *8*(6), 1697-1703.

Roost, M. S., Slieker, R. C., Bialecka, M., Van Iperen, L., Gomes Fernandes, M. M., He, N., ... & Chuva de Sousa Lopes, S. M. (2017). DNA methylation and transcriptional trajectories during human development and reprogramming of isogenic pluripotent stem cells. *Nature Communications*, *8*(1), 908.

Papp, B., & Plath, K. (2013). Epigenetics of reprogramming to induced pluripotency. *Cell*, *152*(6), 1324-1343.

Shi, Y., Wang, M., Mi, D., Lu, T., Wang, B., Dong, H., ... & Wang, X. (2021). Mouse and human share conserved transcriptional programs for interneuron development. *Science*, *374*(6573), eabj6641.

Polo, J. M., Liu, S., Figueroa, M. E., Kulalert, W., Eminli, S., Tan, K. Y., ... & Hochedlinger, K. (2010). Cell type of origin influences the molecular and functional properties of mouse induced pluripotent stem cells. *Nature biotechnology*, *28*(8), 848-855.

Ghosh, Z., Wilson, K. D., Wu, Y., Hu, S., Quertermous, T., & Wu, J. C. (2010). Persistent donor cell gene expression among human induced pluripotent stem cells contributes to differences with human embryonic stem cells. *PloS one*, *5*(2), e8975.

Bar-Nur, O., Russ, H. A., Efrat, S., & Benvenisty, N. (2011). Epigenetic memory and preferential lineage-specific differentiation in induced pluripotent stem cells derived from human pancreatic islet beta cells. *Cell stem cell*, *9*(1), 17-23.

Xu, H., Yi, B. A., Wu, H., Bock, C., Gu, H., Lui, K. O., ... & Chien, K. R. (2012). Highly efficient derivation of ventricular cardiomyocytes from induced pluripotent stem cells with a distinct epigenetic signature. *Cell research*, *22*(1), 142-154.

Khan, N. M., Diaz-Hernandez, M. E., Chihab, S., Priyadarshani, P., Bhattaram, P., Mortensen, L. J., ... & Drissi, H. (2023). Differential chondrogenic differentiation between iPSC derived from healthy and OA cartilage is associated with changes in epigenetic regulation and metabolic transcriptomic signatures. *Elife*, *12*, e83138.

Morris, T. J., Butcher, L. M., Feber, A., Teschendorff, A. E., Chakravarthy, A. R., Wojdacz, T. K., & Beck, S. (2014). ChAMP: 450k chip analysis methylation pipeline. *Bioinformatics*, *30*(3), 428-430.

Aryee, M. J., Jaffe, A. E., Corrada-Bravo, H., Ladd-Acosta, C., Feinberg, A. P., Hansen, K. D., & Irizarry, R. A. (2014). Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics*, *30*(10), 1363-1369.

Fortin, J. P., Triche Jr, T. J., & Hansen, K. D. (2017). Preprocessing, normalization and integration of the Illumina HumanMethylationEPIC array with minfi. *Bioinformatics*, *33*(4), 558-560.

Zhou, W., Laird, P. W., & Shen, H. (2017). Comprehensive characterization, annotation and innovative use of Infinium DNA methylation BeadChip probes. *Nucleic acids research*, *45*(4), e22-e22.

Teschendorff, A. E., Marabita, F., Lechner, M., Bartlett, T., Tegner, J., Gomez-Cabrero, D., & Beck, S. (2013). A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics*, *29*(2), 189-196.

Smyth, G. K. (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Statistical applications in genetics and molecular biology*, *3*(1).

Wettenhall, J. M., & Smyth, G. K. (2004). limmaGUI: a graphical user interface for linear modeling of microarray data. *Bioinformatics*, *20*(18), 3705-3706.

Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W. M., Zheng, S., Butler, A., ... & Satija, R. (2021). Integrated analysis of multimodal single-cell data. *Cell*, *184*(13), 3573-3587.

Hao, Y., Stuart, T., Kowalski, M. H., Choudhary, S., Hoffman, P., Hartman, A., ... & Satija, R. (2024). Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nature biotechnology*, *42*(2), 293-304.

Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., ... & Rinn, J. L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nature biotechnology*, *32*(4), 381-386.

Qiu, X., Mao, Q., Tang, Y., Wang, L., Chawla, R., Pliner, H. A., & Trapnell, C. (2017). Reversed graph embedding resolves complex single-cell trajectories. *Nature methods*, *14*(10), 979-982.

Cao, J., Spielmann, M., Qiu, X., Huang, X., Ibrahim, D. M., Hill, A. J., ... & Shendure, J. (2019). The single-cell transcriptional landscape of mammalian organogenesis. *Nature*, *566*(7745), 496-502.

Chen, E. Y., Tan, C. M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G. V., ... & Ma’ayan, A. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC bioinformatics*, *14*, 1-14.

Kuleshov, M. V., Jones, M. R., Rouillard, A. D., Fernandez, N. F., Duan, Q., Wang, Z., ... & Ma'ayan, A. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic acids research*, *44*(W1), W90-W97.

Xie, Z., Bailey, A., Kuleshov, M. V., Clarke, D. J., Evangelista, J. E., Jenkins, S. L., ... & Ma'ayan, A. (2021). Gene set knowledge discovery with Enrichr. *Current protocols*, *1*(3), e90.

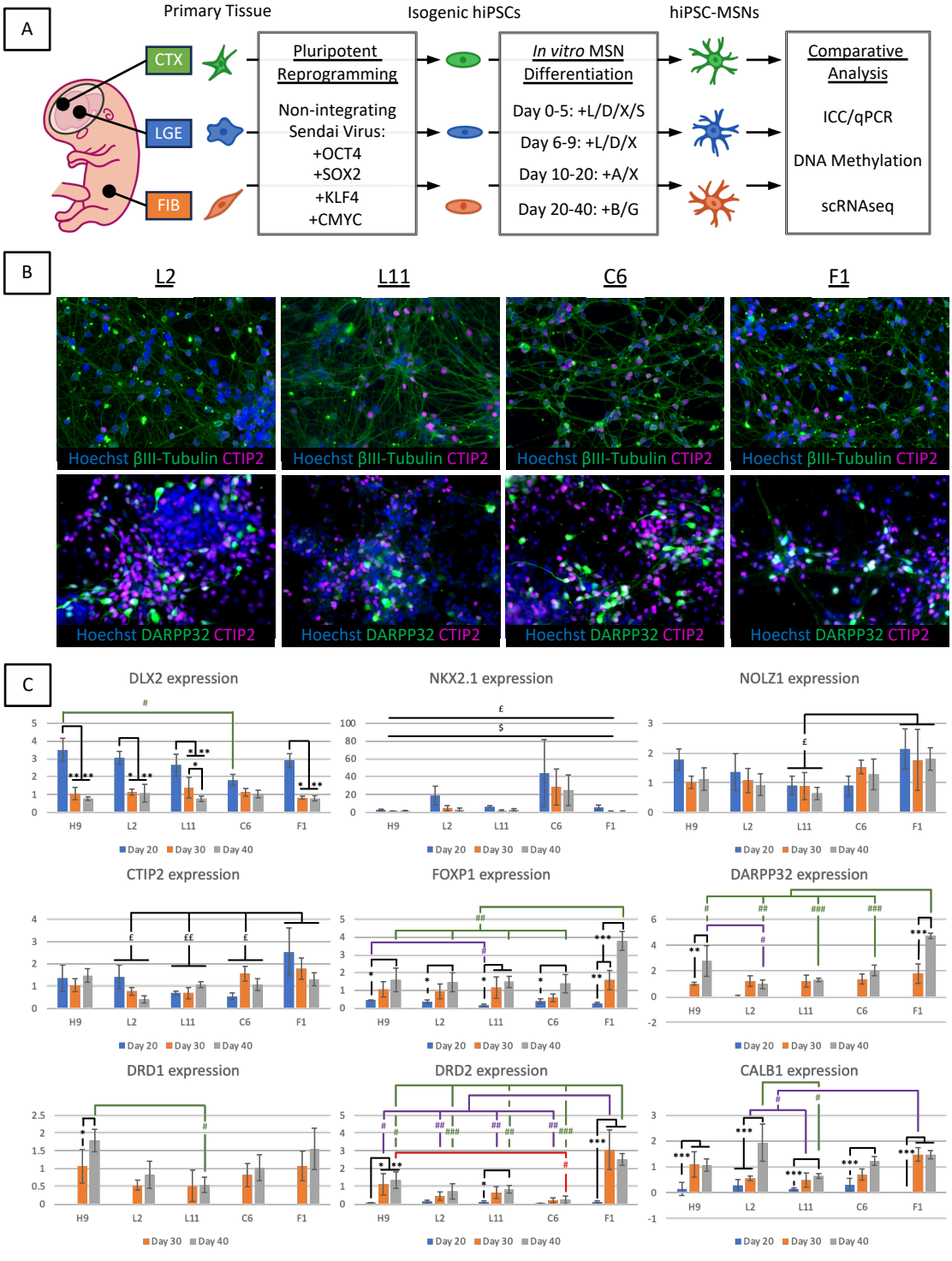



**Figure 1** – Isogenic hiPSCs exhibit variable capacity to differentiate into hPSC-MSNs

**A.** Study overview. CTX = cortex; FIB = fibroblasts; L = LDN193189; D = dorsomorphin; X = XAV939; S = SB431542; A = ActivinA; B = BDNF; G = GDNF; ICC = immunocytochemistry. **B.** Immunocytochemical staining in L2, L11, C6, and F1 for β-III Tubulin (green, top), DARPP32 (green, bottom), CTIP2 (magenta, both) and Hoechst (blue, both) at DIV 30 (top) and DIV 40 (bottom) of MSN differentiation. **C.** Gene expression analysis using qPCR on RNA collected from L2, L11, C6, F1 and H9 derived hPSC-MSNs at DIV 20, 30 and 40 of independent MSN differentiations (n=3 differentiations per cell line). Data is normalised to DIV 30 H9 derived MSNs. Significant results are depicted on bar charts where £ symbolises main effects of line, $ symbolises main effects of Day, # symbolises interactions between Line/Day and * symbolises interactions between Day/Line. The number of symbols indicates p value, where 1 symbol = $p<0.05$, 2 = $p<0.01$, and 3 = $p<0.001$.

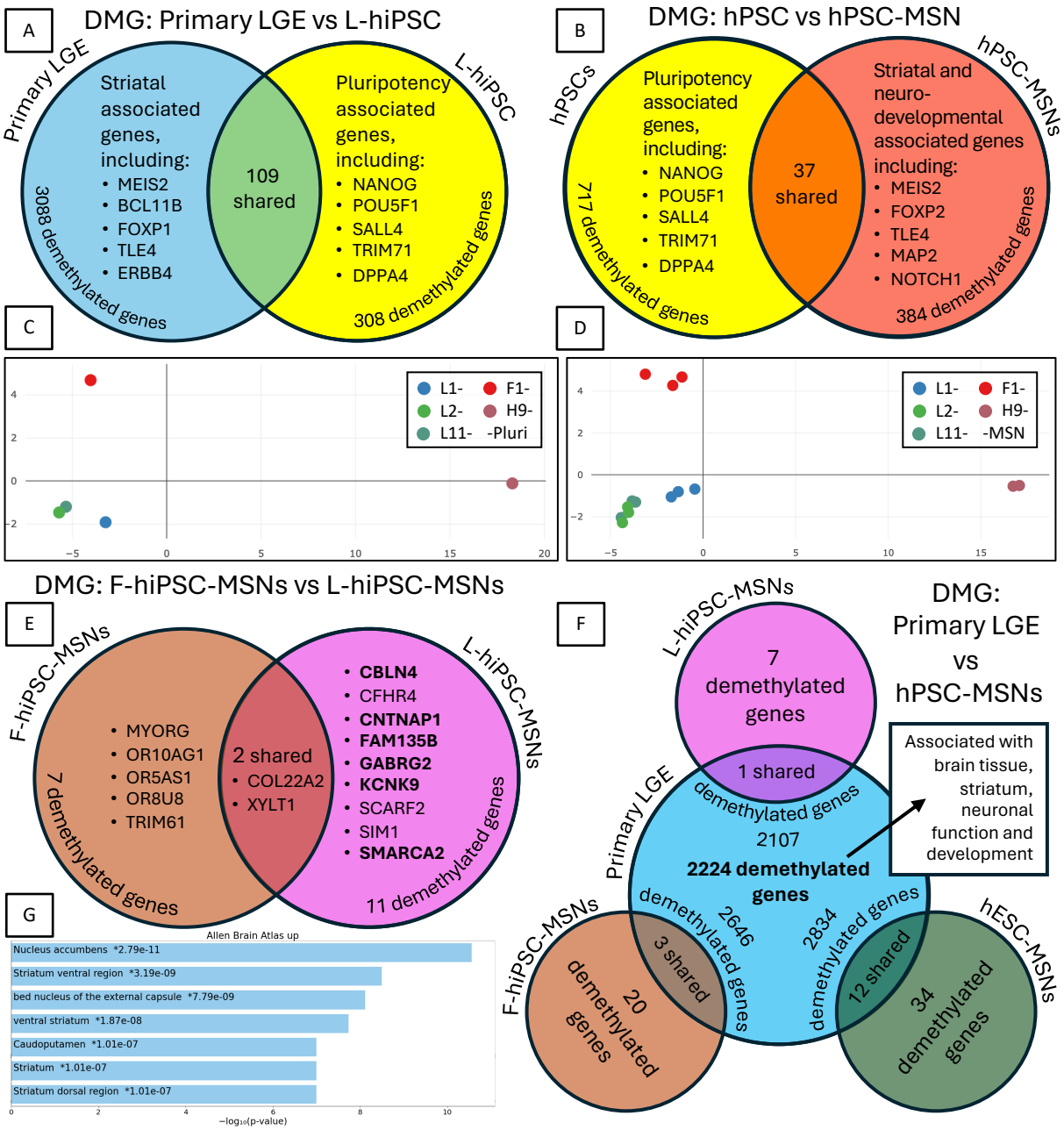



**Figure 2** – DNA methylation analysis of hPSCs, hPSC-MSNs and primary LGE

**A-B & E-F.** Venn diagrams summarising significant differentially methylated genes (DMG), as demethylated genes. DMG comparisons are between (A) primary LGE (left, blue) and pluripotent L-hiPSCs (right, yellow); (B) pluripotent hPSCs (left, yellow) and differentiated hPSC-MSNs (right, orange); (E) F-hiPSC-MSNs (left, tan) and L-hiPSC-MSNs (right, pink); (F) Primary LGE (centre, blue), L-hiPSC-MSNs (top centre, pink), F-hiPSC-MSNs (bottom left, tan), and hESC-MSNs (bottom right, green). **C-D.** MDS plots of (C) pluripotent hPSCs and (D) hPSC-MSNs, produced using top 1000 most variable CPGs. **G.** Bar chart of the most significantly enriched terms from the Allen brain atlas up database for the 2224 genes that were significantly demethylated in Primary LGE compared to hPSC-MSNs shown in F.

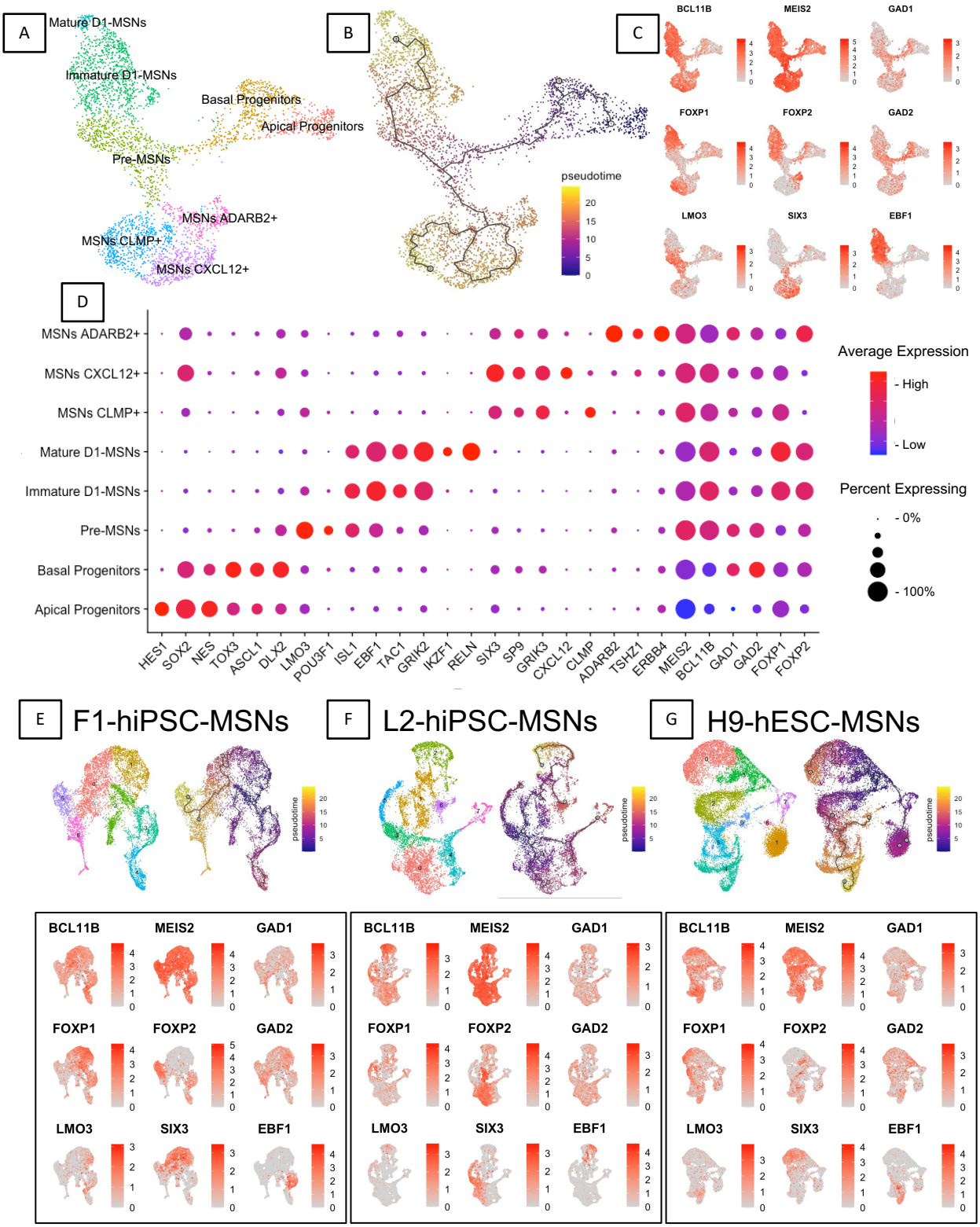
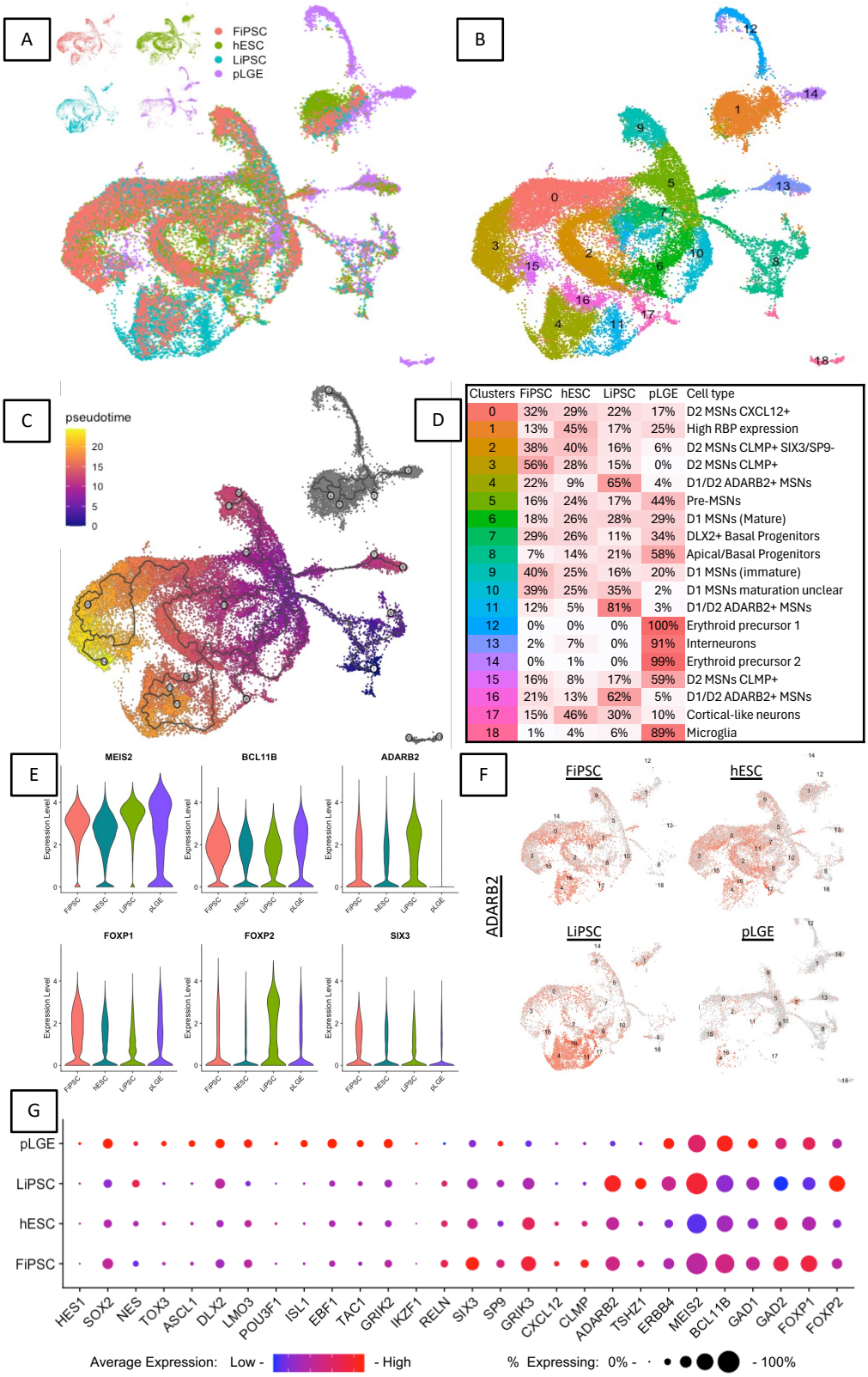



**Figure 3** – Individual scRNAseq analysis of MSNs derived from authentic LGE and hPSC-MSNs

**A.** UMAP of primary LGE (pLGE) derived cells of MSN lineage, annotated using canonical gene expression. **B.** Pseudotime analysis trajectory plot mapped onto pLGE UMAP depicted in A, showing predicted developmental path from apical progenitors to maturing MSNs with a major fate choice in the pre-MSN stage. **C.** Expression plots of key MSN markers in pLGE. **D.** Dotplot showing expression of key MSN markers across pLGE clusters, coloured according to average expression across a cluster's population and percentage of cluster expressing the gene. **E-G.** Summary UMAP plots showing clusters, pseudotime trajectory, and expression of key MSN markers in hiPSC-MSNs, expanded in Figures S6-8.




| Clusters | FiPSC | hESC | LiPSC | pLGE | Cell type |
|---|---|---|---|---|---|
| 0 | 32% | 29% | 22% | 17% | D2 MSNs CXCL12+ |
| 1 | 13% | 45% | 17% | 25% | High RBP expression |
| 2 | 38% | 40% | 16% | 6% | D2 MSNs CLMP+ SIX3/SP9- |
| 3 | 56% | 28% | 15% | 0% | D2 MSNs CLMP+ |
| 4 | 22% | 9% | 65% | 4% | D1/D2 ADARB2+ MSNs |
| 5 | 16% | 24% | 17% | 44% | Pre-MSNs |
| 6 | 18% | 26% | 28% | 29% | D1 MSNs (Mature) |
| 7 | 29% | 26% | 11% | 34% | DLX2+ Basal Progenitors |
| 8 | 7% | 14% | 21% | 58% | Apical/Basal Progenitors |
| 9 | 40% | 25% | 16% | 20% | D1 MSNs (immature) |
| 10 | 39% | 25% | 35% | 2% | D1 MSNs maturation unclear |
| 11 | 12% | 5% | 81% | 3% | D1/D2 ADARB2+ MSNs |
| 12 | 0% | 0% | 0% | 100% | Erythroid precursor 1 |
| 13 | 2% | 7% | 0% | 91% | Interneurons |
| 14 | 0% | 1% | 0% | 99% | Erythroid precursor 2 |
| 15 | 16% | 8% | 17% | 59% | D2 MSNs CLMP+ |
| 16 | 21% | 13% | 62% | 5% | D1/D2 ADARB2+ MSNs |
| 17 | 15% | 46% | 30% | 10% | Cortical-like neurons |
| 18 | 1% | 4% | 6% | 89% | Microglia |

**Figure 4** – Integrated scRNAseq analysis of MSNs derived from authentic LGE and hPSC-MSNs

**A.** UMAPs of integrated data set, showing overlap between cell populations (main), and individual maps of each sample (small). **B.** Cell clusters depicted on UMAP plot. **C.** Pseudotime analysis trajectory plot mapped onto UMAP. **D.** Table depicting each cluster's predicted phenotype, and the contributing proportion by sample. **E.** Violin plots showing differences in expression level across total populations for differentially expressed striatal genes. **F.** Expression plots for ADARB2 split by sample **G.** Dotplot showing expression of key MSN markers across samples, coloured according to average expression across a sample's population and sized by percentage of cells expressing the gene.