Citation: Molecular Therapy — Methods & Clinical Development (2015) **2,** 15030; doi:10.1038/mtm.2015.30
Official journal of the American Society of Gene & Cell Therapy

www.nature.com/mtm

ARTICLE

Quantitative high-throughput gene expression profiling of human striatal development to screen stem cell–derived medium spiny neurons

Marco Straccia[1], Gerardo Garcia-Diaz Barriga[1], Phil Sanders[1], Georgina Bombau[1], Jordi Carrere[1], Pedro Belio Mairal[1], Ngoc-Nga Vinh[2], Sun Yung[2], Claire M Kelly[2], Clive N Svendsen[3], Paul J Kemp[2], Jamshid Arjomand[4,5], Ryan C Schoenfeld[4,6], Jordi Alberch[1], Nicholas D Allen[2], Anne E Rosser[2] and Josep M Canals[1]

A systematic characterization of the spatio-temporal gene expression during human neurodevelopment is essential to understand brain function in both physiological and pathological conditions. In recent years, stem cell technology has provided an *in vitro* tool to recapitulate human development, permitting also the generation of human models for many diseases. The correct differentiation of human pluripotent stem cell (hPSC) into specific cell types should be evaluated by comparison with specific cells/tissue profiles from the equivalent adult *in vivo* organ. Here, we define by a quantitative high-throughput gene expression analysis the subset of specific genes of the whole ganglionic eminence (WGE) and adult human striatum. Our results demonstrate that not only the number of specific genes is crucial but also their relative expression levels between brain areas. We next used these gene profiles to characterize the differentiation of hPSCs. Our findings demonstrate a temporal progression of gene expression during striatal differentiation of hPSCs from a WGE toward an adult striatum identity. Present results establish a gene expression profile to qualitatively and quantitatively evaluate the telencephalic hPSC-derived progenitors eventually used for transplantation and mature striatal neurons for disease modeling and drug-screening.

Molecular Therapy — Methods & Clinical Development (2015) **2**, 15030; doi:10.1038/mtm.2015.30; published online 16 September 2015

INTRODUCTION

The study of human neurodevelopment is essential to understand the physiological function in both normal brain development and during disease processes. The important discovery that induced pluripotent stem cells (iPSCs) can be generated from human somatic cells[1] allows the study human development and modeling of human diseases *in vitro* using iPSCs derived from patients and healthy individuals. However, the challenge is to efficiently and correctly differentiate human iPSCs to the desired cell type or types. Many human stem cell differentiation protocols are based on mouse developmental data despite the various differences that exist between mouse and human development. The ideal method to differentiate iPSCs would be to specifically recapitulate human development *in vitro* by activating, inhibiting, and tuning specific biochemical pathways in the correct temporal manner. In addition, many human fetal studies do not analyze and/or compare the levels of expression with the adulthood, missing critical information about the relative expression levels.

Gene expression profiling during human development is crucial for defining stage-related changes for specific anatomical regions, and these data could subsequently be applied to *in vitro* differentiation protocols. In the last decade, many transcriptomic approaches have been used to analyze gene expression in the human brain.[2,3] While these techniques produce large amounts of data, they do not provide quantitative gene expression information and the results that are obtained require validation. Quantitative real-time polymerase chain reaction (qRT-PCR) is the gold standard for producing quantitative and validated gene expression data.

For neurodegenerative diseases, differentiation of human pluripotent stem cell (hPSC) into neural cells should be evaluated by comparison with specific profiles from the equivalent *in vivo* nervous system area. For example, the generation of hPSC-derived medium spiny neurons (MSNs) to study Huntington's disease is on the state of the art. In fact several efforts have been made to generate human PSC-derived MSNs with varying degrees of success.[4–7] However, while some papers compare to the fetal gene expression

[1]Department of Cell Biology, Immunology and Neuroscience, Faculty of Medicine, August Pi i Sunyer Biomedical Research Institute (IDIBAPS), and Networked Biomedical Research Centre for NeuroDegenerative Disorders (CIBERNED), University of Barcelona, Barcelona, Spain; [2]Cardiff Repair Group, School of Biosciences and Medicine, Cardiff University, Cardiff, UK; [3]Regenerative Medicine Institute, Cedars-Sinai Medical Center, Los Angeles, California, USA; [4]CHDI Foundation, Princeton, New Jersey, USA; [5]Present address: Genea Biocells, Princeton, New Jersey, USA; [6]Present address: Janssen Research & Development, Pharmaceutical Companies of Johnson & Johnson, Spring House, Pennsylvania, USA. Correspondence: JM Canals (jmcanals@ub.edu)
Received 9 July 2015; accepted 22 July 2015

profiles of the whole ganglionic eminence (WGE; the striatal primordium),[4] none of them compare to the adult caudate-putamen profile. Therefore, it is necessary to further characterize the WGE human development and its relative levels with the adult caudate-putamen in order to efficiently and correctly differentiates human iPSCs to MSNs.

In the present work, we focus on the generation of quantitative genetic profiles in human fetal and adult brain by low-cost high-throughput qRT-PCR. We identified specific genes and their expression levels that allow determining the developmental stage of the hPSC-derived striatal cells. Our results provide a detailed dataset of genes involved in striatal development that can be used to assess the quality and efficiency of current and future protocols that differentiate hPSCs to striatal neurons, thereby improving and refining MSN differentiation protocols for cell transplantation, disease modeling, and drug-screening.

RESULTS

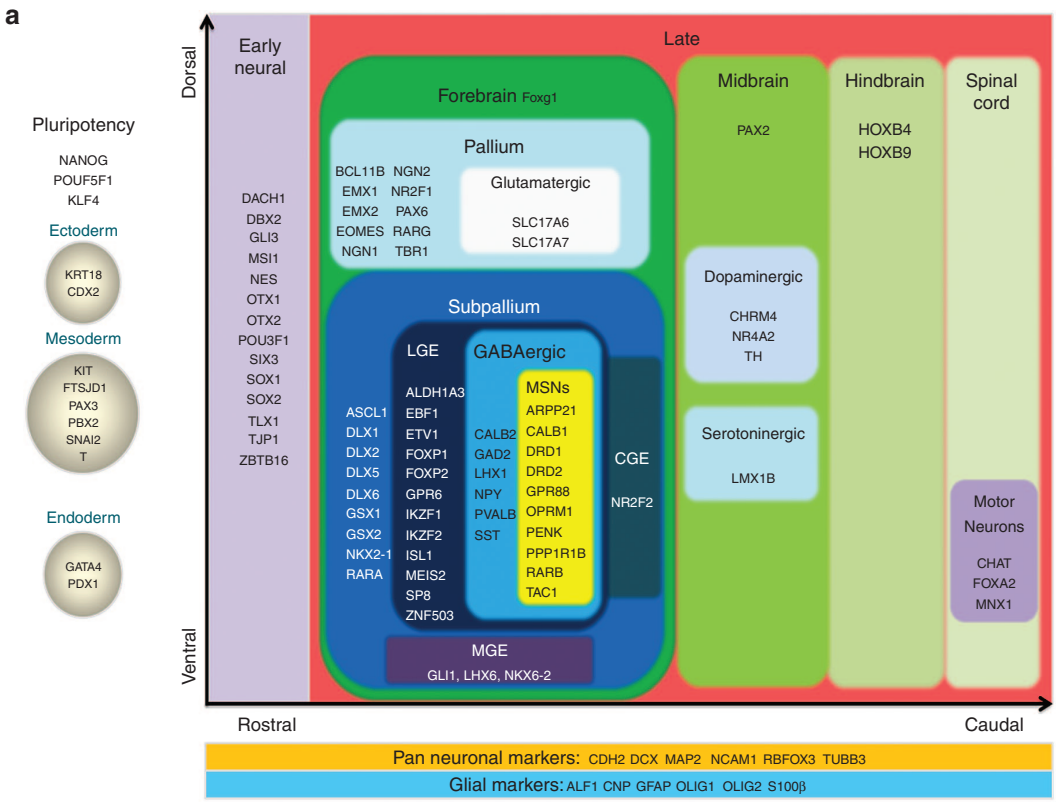
A selected gene set distinguishes human subpallial derivatives from fetal pallium and adult motor cortex during development

We defined a panel of 106 genes (Figure 1a and Table 1) that are known from the literature to be involved in mammalian brain development, with an enrichment for genes implicated in subpallial specification. Expression of this gene set in dissected human brain samples was quantitatively analyzed using OpenArray nanoscale real-time PCR technology. Samples included WGE and cortex dissected from fetal tissues (between 49 and 63 pcd) as well as adult caudate, putamen, and motor cortex (see Supplementary Figure S1a for complete details of tissue sources and specimens).

Unbiased hierarchical clustering of gene expression data (Figure 1b) from the 106 genes grouped human samples according to their developmental stage. Subclusters A, B, and C identified the adult samples with subcluster B containing the adult motor cortex samples and subcluster C containing all the striatal samples. Interestingly, the caudate and putamen samples clustered together and a gene-by-gene analysis did not identify any statistical differences between them. This indicates that our gene set does not distinguish between these two nuclei of the adult striatum. For this reason, the samples from both regions were combined and referred to as striatum. Subcluster A clustered outside of the main group analysis indicating a distinct gene expression pattern. Since this subcluster was mainly composed of samples dissected from the same brain (both cortical and striatal tissues), this suggests a donor-related issue, inaccurate dissection, or poor postmortem preservation. However, we could not exclude these samples from our whole gene set analysis on a statistical basis, since they were not identified as biological group outliers by the multivariate Student's *t*-test. Gene expression differences between striatum and cortex are clearly evidenced on the heat map.

The fetal samples were clustered into two subclusters (Figure 1b, groups D and E). Subcluster D contained all the WGE samples and subcluster E contained the fetal cortical samples. Again, gene expression differences between WGE and cortical primordia are clearly shown in the heat map (Figure 1b).

Having determined that our customized gene expression assay can identify different human brain samples as distinct biological groups, we were then interested in identifying those genes that specifically characterized each biological group. To identify groups of genes that are specifically associated with each sample cluster, we performed a k-means analysis on the gene expression data (Figure 2). This approach generated three clusters for the fetal samples (clusters 2, 4, and 5; Figure 2a,b). Clusters 2 (27 genes) and 5 (9 genes) contained genes that are expressed in both fetal cortex and WGE (Figure 2b). Cluster 4 (14 genes) contained those genes that result specific for WGE (Figure 2b).



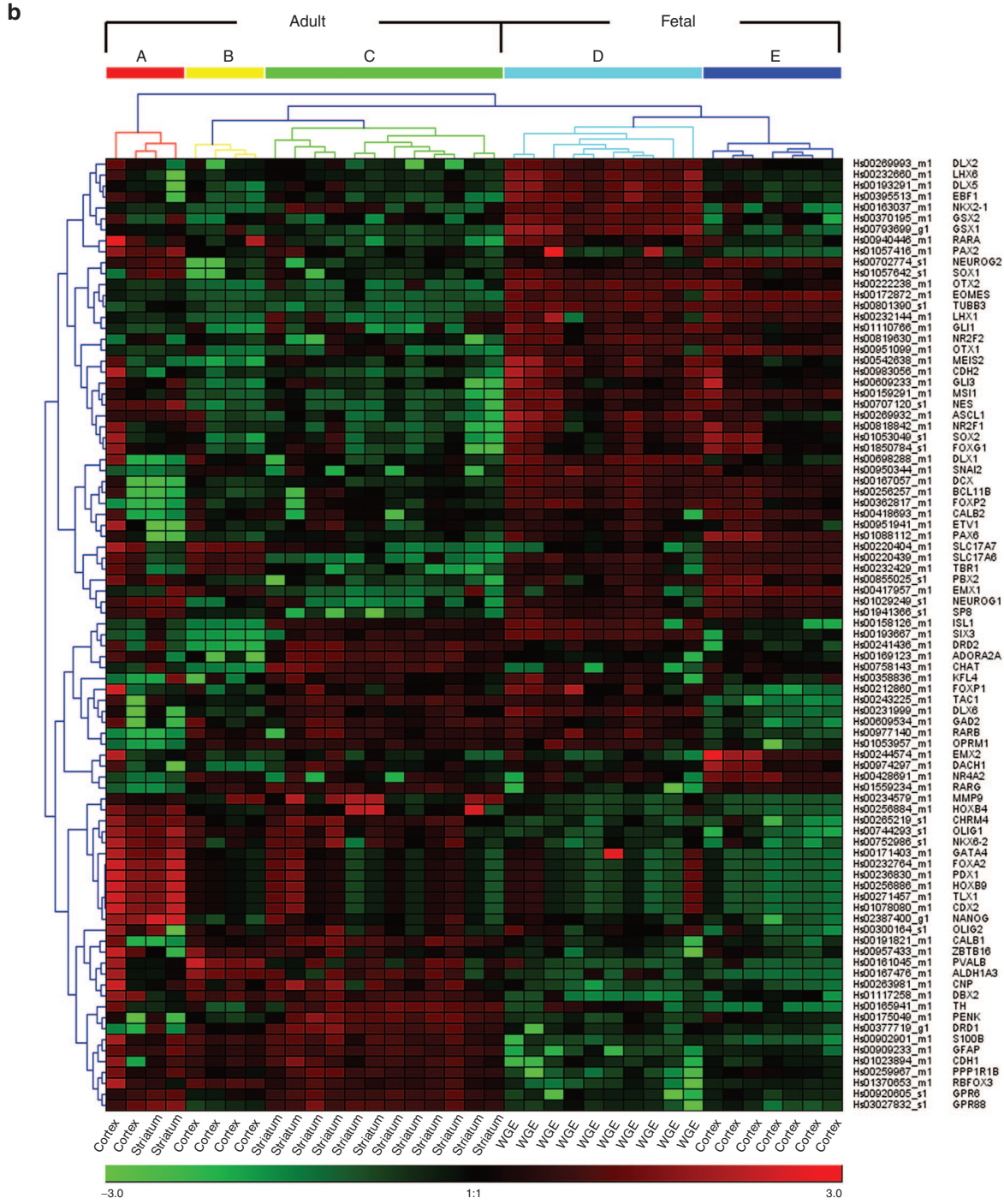


Figure 1 Gene expression view of telencephalic development. (**a**) Developmental scheme for the 106 genes selected for quantitative gene expression analysis in human adult and fetal forebrain pallial and subpallial derivatives. Panel shows the selected gene set as markers during mammalian central nervous system (CNS) development. Genes to monitor pluripotency and to determine germ layer identity are represented outside of the Cartesian diagram. Inside the diagram, the genes are grouped into colored boxes. The x-axis represents the Rostro-Caudal axis of human brain development. According to these criteria, the genes are located depending mainly on their expression during CNS development. Early neural genes are shown on the left side of the diagram. Below the diagram general neuronal and glial gene markers are indicated. Subpallial genes are over-represented in this gene set since we focus on striatal development. (**b**) Hierarchical clustering with complete linkage clustering groups the samples into five clusters. Subcluster A (red) contains a mix of adult cortical and caudate-putamen nuclei samples, suggesting technical bias (see text for details). Subcluster B (yellow) corresponds to adult cortical samples. Subcluster C (green) contains solely adult striatal samples. Subclusters D (light blue) and E (dark blue) contain all of the fetal samples with D containing the whole ganglionic eminence (WGE) samples and E containing the cortical tissue samples. Colors range from −3.0 green to +3.0 red depending on the double ΔC_t of each sample/gene compared to the average double ΔC_t (black, 1:1) of all of the expressed genes.

Table 1 Selected gene set to assess expression during central nervous system (CNS) development

Group		Genes
A	Pluripotency	*KIT, KFL4, MSI1, NANOG, POU5F1,*
B	Germ layers	*CDX2, FTSJD1, GATA4, KRT18, PAX3, PBX2, PDX1, SNAI2, T*
C	Early neural	*DACH1, DBX2, NES, OTX1, OTX2, POU3F1, SIX3, SOX2, SOX1, TLX1, ZBTB16*
D	Forebrain	*FOXG1*
E	Pallium	*EOMES, EMX2, EMX1, NEUROG1, NEUROG2, NR2F1, PAX6, SLC17A7, SLC17A6, TBR1,*
F	Subpallium	*ASCL1, DLX1, DLX2, DLX5, DLX6, GSX1, GSX2, NKX2-1, SP8*
G	LGE	*ALDH1A3, BCL11B, EBF1, ETV1, FOXP1, FOXP2, GPR6, IKZF1, IKZF2, ISL1, MEIS2, RARB, ZNF503*
H	GABAergic	*CALB2, GAD2, LHX1, NPY, PVALB, SST*
I	Medium spiny neurons	*ADORA2A, ARPP21, CALB1, DRD1, DRD2, GPR88, OPRM1, PENK, PPP1R1B, TAC1*
J	MGE, CGE	*GLI1, LHX6, NKX6-2, NR2F2*
K	Midbrain	*CHRM4, LMX1B, NR4A2, PAX2, TH*
L	Hindbrain	*HOXB4, HOXB9*
M	Spinal cord	*CHAT, FOXA2, MNX1*
N	Neuronal marker	*CDH2, DCX, MAP2, NCAM1, RBFOX3, TUBB3*
O	Glial markers	*AIF1, CNP, GFAP, OLIG1, OLIG2, S100B*
P	Others	*CDH1, GLI3, MMP9, RARA, RARG, TJP1*

Genes are grouped based on their spatio-temporal biological relevance during CNS development according to the currently available scientific literature. The letters in the first column ("Group") identify the groups of genes to interpret the mRNA expression profile graph in Figures 3a and 4a.
CGE, caudal ganglionic eminence; LGE, lateral ganglionic eminence; MGE, medial ganglionic eminence.

K-means genes clustering of adult cortical and striatal samples produced four clusters (clusters 1, 3, 6, and 7; Figure 2a,c). Cluster 1 contained nine genes (*CDX2, FOXA2, GATA4, HOXB9, NANOG, NKX6-2, PAX2, PDX1, and TLX1*) that were expressed in cluster A samples (Figure 1b), with these samples showing unexpected higher expression of certain genes that was not present in the other tissue donors. Clusters 3 (13 genes) and 7 (3 genes) contained genes that are commonly expressed in the adult motor cortex, putamen, and caudate (Figure 2c), while Cluster 6 contained 12 genes that are highly expressed in caudate and putamen, identifying adult striatum (Figure 2c).

Thus using these approaches, we have demonstrated that our quantitative gene expression assay can generate subsets of genes that identify brain subregions at specific developmental stages.

Quantitative expression of a discrete number of genes distinguishes WGE from fetal cortex

After rostral-caudal patterning *in vivo*, progenitor cells must commit to a dorsal or ventral fate.[8,9] Similarly, stem cells differentiating toward striatal neurons must be directed to a subpallial (ventral telencephalon) fate following telencephalic commitment for the subsequent specification to MSNs. With this in mind, it is important to determine the gene expression signature that distinguishes human fetal cortex from WGE. We compared the gene expression profiles that we obtained from the fetal cortex and WGE samples (7–9 post-conception weeks (p.c.w.)) with their total gene expression signatures being represented by a spike graph (Figure 3a) to facilitate the visual comparison with the 106 genes being grouped according to biological relevance (Table 1).

Comparison of the fetal cortex and WGE gene expression data revealed opposing gene expression signatures between the two areas (Figure 3a). The WGE was characterized by high expression of the early neural gene *SIX3* and genes involved in lateral ganglionic eminence (LGE), medial ganglionic eminence, and caudal ganglionic eminence specification (Figure 3a, groups F, G, I, and J), while the fetal cortex showed increased expression of *OTX1* and *DACH1* and pallial markers (Figure 3a, groups D and E). Furthermore, expression of calretinin (*CALB2*) and *PENK* was also observed in fetal cortex.

A set of 19 statistically significant differentially expressed genes distinguished WGE from fetal cortex. Figure 3b shows the mRNA expression levels of the 14 genes that were upregulated (red bars) and the 5 that were downregulated (green bars) in WGE compared to fetal cortex. Interestingly, at this stage of human forebrain development *DARPP-32* (*PPP1R1B*), which is considered a specific marker for mature MSNs,[5,10] is more highly expressed in the cortical region than in the WGE. *NANOG* expression was significantly increased in WGE, while significantly higher *RBFOX3* (*NeuN*) expression was observed in fetal cortex (see also low expressed genes in Supplementary Figure S2).

The statistical analysis of the specific marker genes for WGE and fetal cortex identified *TAC1*, *DLX5,* and *DLX6* (P value < 0.0001) together with *DLX1*, *DLX2*, *EBF1,* and *LHX6* (P value < 0.01) as the most reliable genetic predictors for human WGE on our panel of selected genes. While the fetal cortex was best identified by two out of the five genes that were downregulated in WGE compared to fetal cortex, namely *TBR1* and *SLC17A7* (P value of 0.01) (Figure 3c). To define the best list of WGE predictor genes, we selected those genes that were upregulated in the WGE compared to the fetal cortex (Figure 3b,c) that were also present in the specific WGE cluster obtained by k-means clustering analysis (Figure 2b; cluster 4). This selection process identified seven genes (*DLX1*, *DLX5*, *DLX6*, *EBF1*, *LHX6*, *NKX2-1,* and *SIX3*), and to ensure that these genes were developmentally specific for WGE, we determined if they were also significantly downregulated in adult striatum. Hence, we compared these seven shortlisted genes with those

genes significantly downregulated in the adult striatum compared to the WGE (Figure 4b,c). This analysis produced a list of gene gene markers for WGE that showed a decrease in gene expression in adult striatum compared to the WGE (*SIX3* 10.5-fold decrease, *LHX6* 11.3-fold decrease, *DLX1* 11.0-fold decrease, *DLX5* 12.7-fold decrease, *DLX6* 2.8-fold decrease, and *EBF1* 12.9-fold decrease). Finally, to establish the statistical significance and the relative quantitative mRNA expression levels of our WGE marker genes among pallial and subpallial derivatives, we performed a gene-by-gene one-way analysis of variance (ANOVA) between WGE, fetal cortex, and adult tissues (striatum and motor cortex) (Figure 3d). Using this approach we have validated six genes whose expression together can be used to identify WGE at this developmental stage.

Identification of a specific gene set to quantitatively distinguish human WGE from adult striatal nuclei

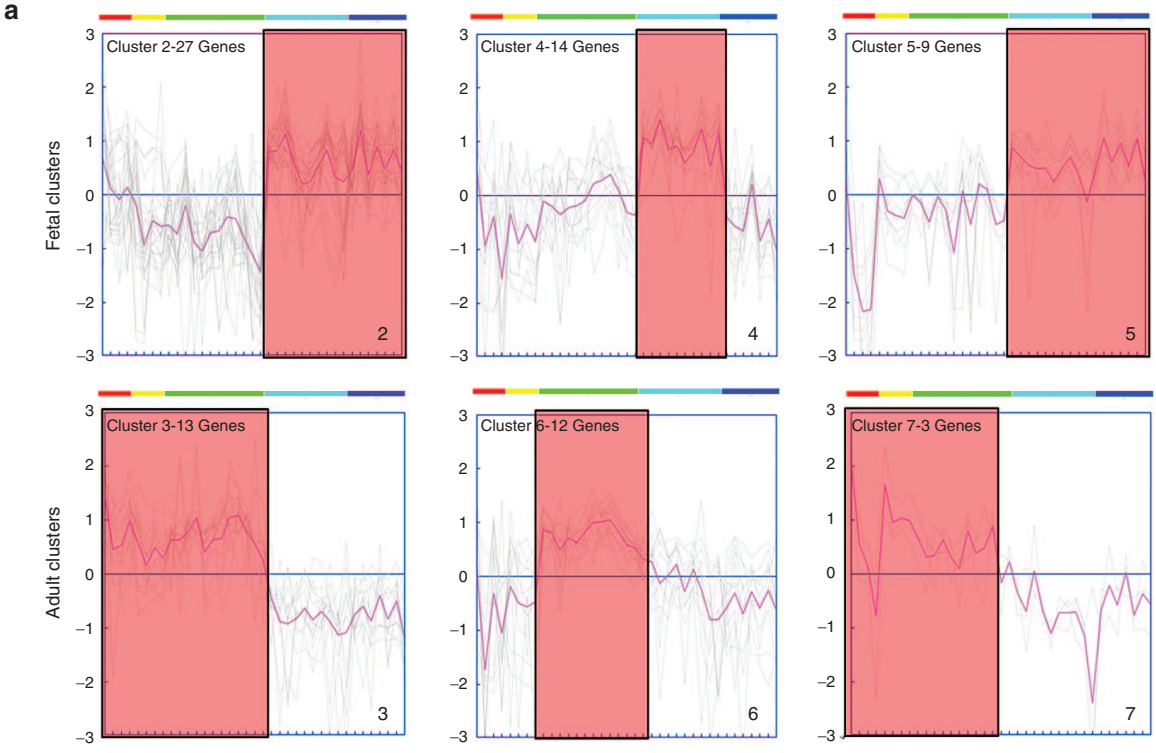
Having identified the critical genes that distinguish human WGE from fetal cortex, next, we characterized the specific quantitative gene expression profile during human striatal development. We studied how the expression of our gene set varied in adult human striatum compared to WGE in order to identify mature striatal gene markers and to observe how these markers change their expression relative to WGE.

An overview of the gene expression data revealed a clear shift of expression from early WGE genes toward mature striatal markers in adult striatum (Figure 4a). A change in neuronal markers' expression was observed between the two stages. Doublecortin (*DCX*), β-III-tubulin (*TUBB3*), and N-cadherin (*CDH2*) were expressed at high levels in the WGE but were downregulated in the striatum, while NeuN (*RBFOX3*) displayed the opposite behavior with its expression positively correlating with tissue maturity (group N, Figure 4a). Glial marker gene expression (*S100B* and *GFAP*, group O, Figure 4a) was higher in adult tissue than in the WGE, indicating the multicellular complexity that is established in the brain parenchyma.

Thirty-seven statistically significant differentially expressed genes defined a quantitative gene expression profile for the adult striatum compared to the WGE (Figure 4b). Eighteen of these genes were more highly expressed in the striatum (Figure 4b, red bars) with 12 of them (*ADORA2A, ARPP21, CHRM4, DRD1, DRD2, CALB1, GPR6, GRP88, PENK, PPP1R1B, TAC1,* and *TH*) being considered from the literature to be specific for striatal neurons (MSNs and interneurons). Furthermore, four oligodendrocyte-associated genes (*S100B, CNP, NKX6-2,* and *OLIG1*) were also significantly upregulated in the striatum. The final two genes that were more highly expressed in the striatum (*RBFOX3* and *RARG*) are associated with the general maturation of brain parenchyma as also shown by the Human Brain Transcriptome project (www.hbatlas.org).[2]

The remaining 19 statistically significant differentially expressed genes were downregulated in the adult human striatum compared to the WGE (Figure 4b, green bars). Early neural genes (*SOX2, MSI1, OTX2,* and *SIX3*), *MASH1* (also known as *ASCL1*) and ganglionic eminence-related genes (*MEIS2, BCL11B, EBF1, LHX6, NR2F2*) were expressed at higher levels in the WGE than in the striatum as were the early neuronal markers *CDH2*, *DCX*, and *TUBB3*. Notably, while the expression of all *DLX* family members (*DLX1, 2, 5,* and *6*) was downregulated compared to striatal nuclei, each family member was downregulated to a different extent. The pallial marker *TBR1* is hardly detectable in adult striatum, while retinoic acid receptor A (*RARA*) expression decreases as development progresses as described elsewhere.[2]

Upregulated genes represented 48.6% of all differentially expressed genes but their average fold change of gene expression was 28.4 compared to an average fold change of 14.7 for the downregulated genes. In addition, the upregulated genes in the adult striatum showed a higher statistical significance ($0 \le P$ value ≤ 0.02) than the downregulated genes ($0 \le P$ value ≤ 0.04). The expression analysis indicated that the most reliable predictor genes ($P < 0.0001$) for striatal maturation were two genes of the G protein-coupled transmembrane receptors family, *GPR6* and *GPR88*, and the adenosine receptor



ADORA2A. Other reliable but less statistically significant predictors for mature striatum were the upregulation of *PPP1R1B, PENK, DRD1, CALB1, DRD2*, and *TAC1* (P value < 0.01), and the downregulation of *BCL11B, MSI1, LHX6, DLX5, DLX1, EBF1, NR2F2*, and *DLX2* (P value < 0.01). Volcano scatter plot analysis (Figure 4d) confirmed the importance of this specific set of striatal upregulated genes as well as the downregulation of subpallium- and WGE-related genes in order to distinguish between these two developmental stages. Thus, we have defined a quantitative gene expression profile for the developmental transition from WGE to adult striatum, which enhances our understanding of the human striatal differentiation process.

Quantitative gene expression signature of human striatal development

Having used quantitative gene expression to define marker gene sets that distinguish between two coexisting developmental areas and two striatal developmental stages, we next identified the

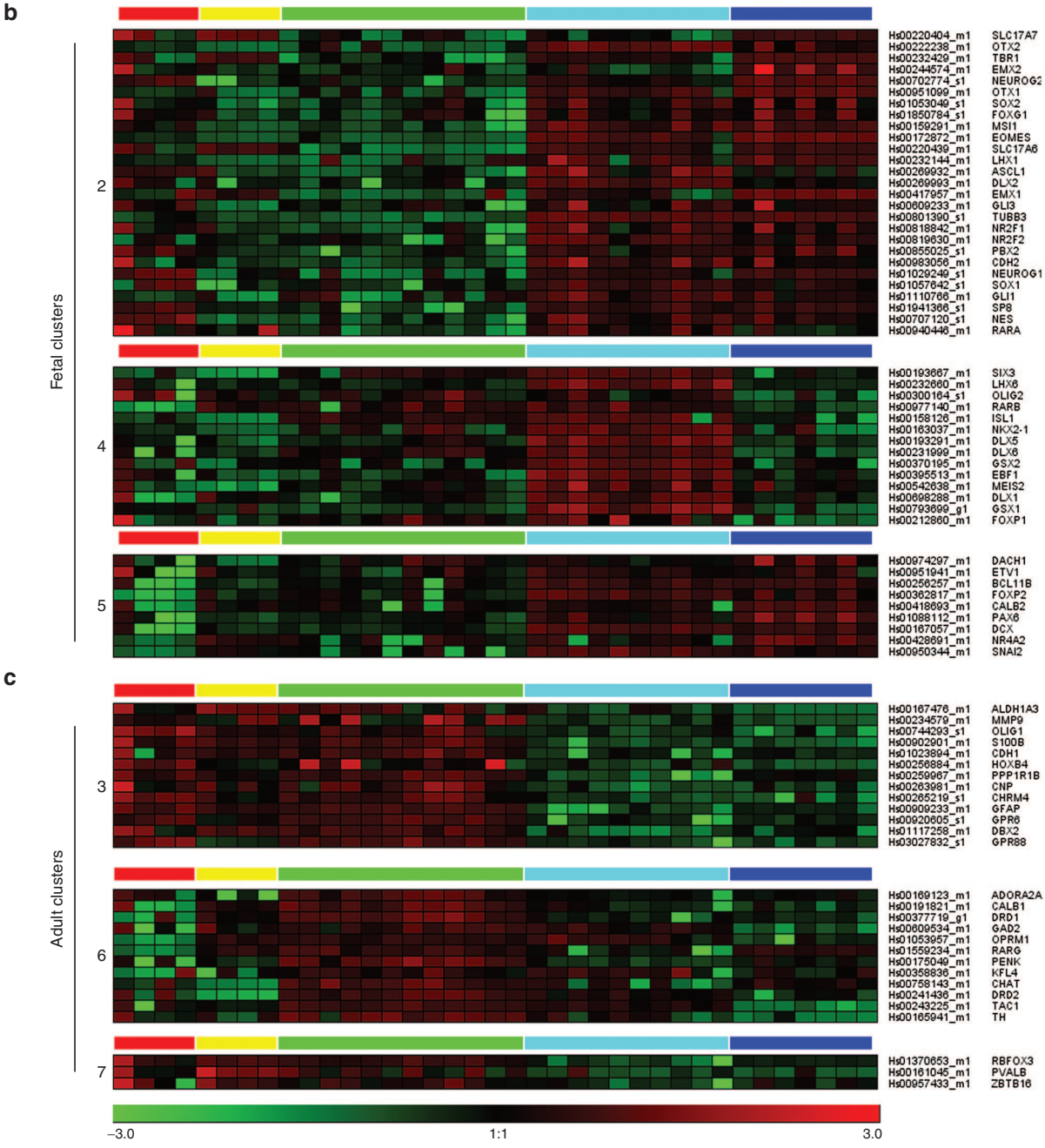
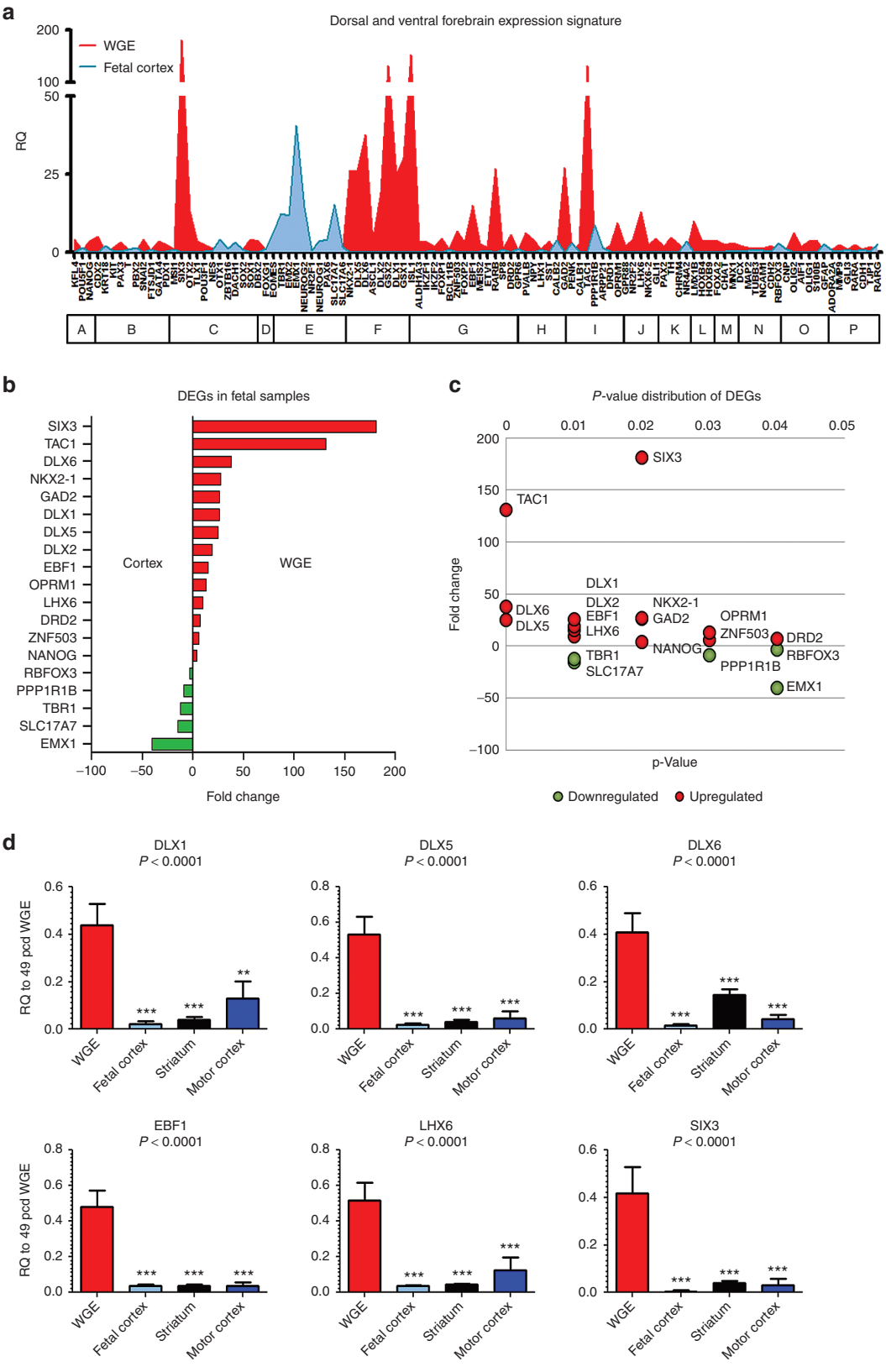


Figure 2 Fetal and adult tissues are distinguished by the expression of specific clusters of genes. (**a**) K-means clustering analysis of the data in Figure 1b generated three fetal-specific clusters (2, 4, and 5) and three adult-specific clusters (3, 6, and 7) indicating genes with higher expression in fetal (cortex and whole ganglionic eminence (WGE)) and adult (motor cortex and caudate-putamen nuclei) human tissue, respectively. Faded red boxes highlight the regions of the expression curves that identify each sample subcluster. (**b**) The heat-maps represent the hierarchical clustering of the genes identified by k-means analysis. Clusters 2 and 5 contain genes commonly increased in fetal WGE and cortex whereas cluster 4 contains a group of genes whose expression is increased exclusively in fetal WGE samples. (**c**) Clusters 3 and 7 contain genes whose expression increased in both adult cortical and striatal samples. Cluster 6 contains genes whose expression is mostly increased in adult striatal samples (Cluster color code is described in Figure 1).



marker genes that can be used to recognize fully mature striatal neurons.

To achieve this, we first considered those genes that were differentially expressed between the WGE and the adult striatum (Figure 4b,c), and then from this gene set, we excluded those genes that were also expressed in the adult cortex (Figure 5a) or have a glial-related function. The genes with increased striatal expression that were identified using this method were only considered as

genuine upregulated striatal marker genes if they were also present in the adult striatum cluster that was generated in the unbiased k-means clustering (Figure 2a,b; cluster 6). This filtering process identified six upregulated striatal markers (Figure 5b; *ADORA2A, CALB1, DRD1, DRD2, PENK,* and *TAC1*). The downregulated striatal markers (Figure 5b) were the six genes (*DLX1, DLX5, DLX6, EBF1, LHX6, and SIX3*) that we previously identified as WGE specific (Figure 3d) and downregulated in adult striatum (Figure 4b,c).

Further to the up- and downregulated striatal markers, we identified two genes (*GAD2* and *OPRM1*) that are expressed at similar levels in the WGE and striatum (*i.e.*, not present in Figure 4b) and they were defined as constant markers (Figure 5b). Finally, we also identified four genes that show a specific decrease in expression in the adult striatum while being consistently expressed in all other tissues examined (*HOXB4, SLC17A6, TBR1,* and *TLX1* (Supplementary Data 2)). These genes were defined as Low-Off striatal markers (Figure 5b). Thus using these criteria, we defined a set of 18 genes that together can be considered as a specific gene expression signature for mature human striatal neurons (Figure 5b).

In order to determine the relative expression level between tissues and the statistical significance of the "up-regulation" and "down-regulation" present within this 18 gene set (Figure 5b), we compared their expression in the striatum to that in the WGE and fetal and adult cortical samples (Figures 3d and 5c).

One-way ANOVA revealed that *ADORA2A*, *CALB1*, *DRD1*, *DRD2, PENK,* and *TAC1* were all reliable markers for mature striatal neurons with *PENK* and *CALB1* expression being 62.3- and 18-fold higher respectively in the striatum compared to the WGE. The other four genes in the upregulated group presented smaller but still statistically significant increases in striatal expression compared to the WGE (*ADORA2A* 14-fold increase, *DRD1* 20.2-fold increase, *DRD2* 6.8-fold increase, and *TAC1* 2.6-fold increase). These findings revealed that direct (*TAC1, DRD1*) and indirect (*PENK, DRD2*) striatal pathway related genes were the most specific markers for adult MSNs together with *CALB1* and *ADORA2A*.

We also analyzed the expression of the *DARPP-32* (*PPP1R1B*) and *CTIP2* (*BCL11B*) genes (Figure 5d) as they are frequently used at the protein level to specifically identify MSNs. Although they were not defined as striatal markers by the k-means clustering analysis described above (Figure 2c; cluster 6), *DARPP-32* (*PPP1R1B*) did display increased expression in the striatum compared to the WGE, however the level of *PPP1R1B* expression that was detected in the human fetal and adult cortex prevented its classification as a striatal specific marker. In fact the resulting ANOVA *P* value for *PPP1R1B* when comparing the striatum to the other tissues tested is 0.0012 (Figure 5d), which is less significant than for any of the six genes we identified as being specifically upregulated in the adult striatum (Figure 5c). *CTIP2* (*BCL11B*) expression decreased 82% in the adult striatum compared to the WGE (Figure 5d). To confirm these results at the protein level, we analyzed CTIP2 protein by western blot in human fetal and adult striatal samples. We observed that although CTIP2 protein was expressed in the oldest of the fetal WGE samples (pcd 61 and 63), no protein was detected in any of the adult striatal samples (Figure 5e). We further analyzed DARPP-32 by immunohistochemistry in human adult brain (Figure 5f). Although human adult putamen showed high number of DARPP-32 positive neuronal soma and projections, cortical white matter and gray matter also contains DARPP-32-expressing cells. In summary, we identified *ADORA2A, CALB1, DRD1, DRD2, PENK*, and *TAC1* as specific adult striatal neuronal marker genes, while *CTIP2* and *DARPP-32* appear not to be reliable markers at the mRNA expression level.

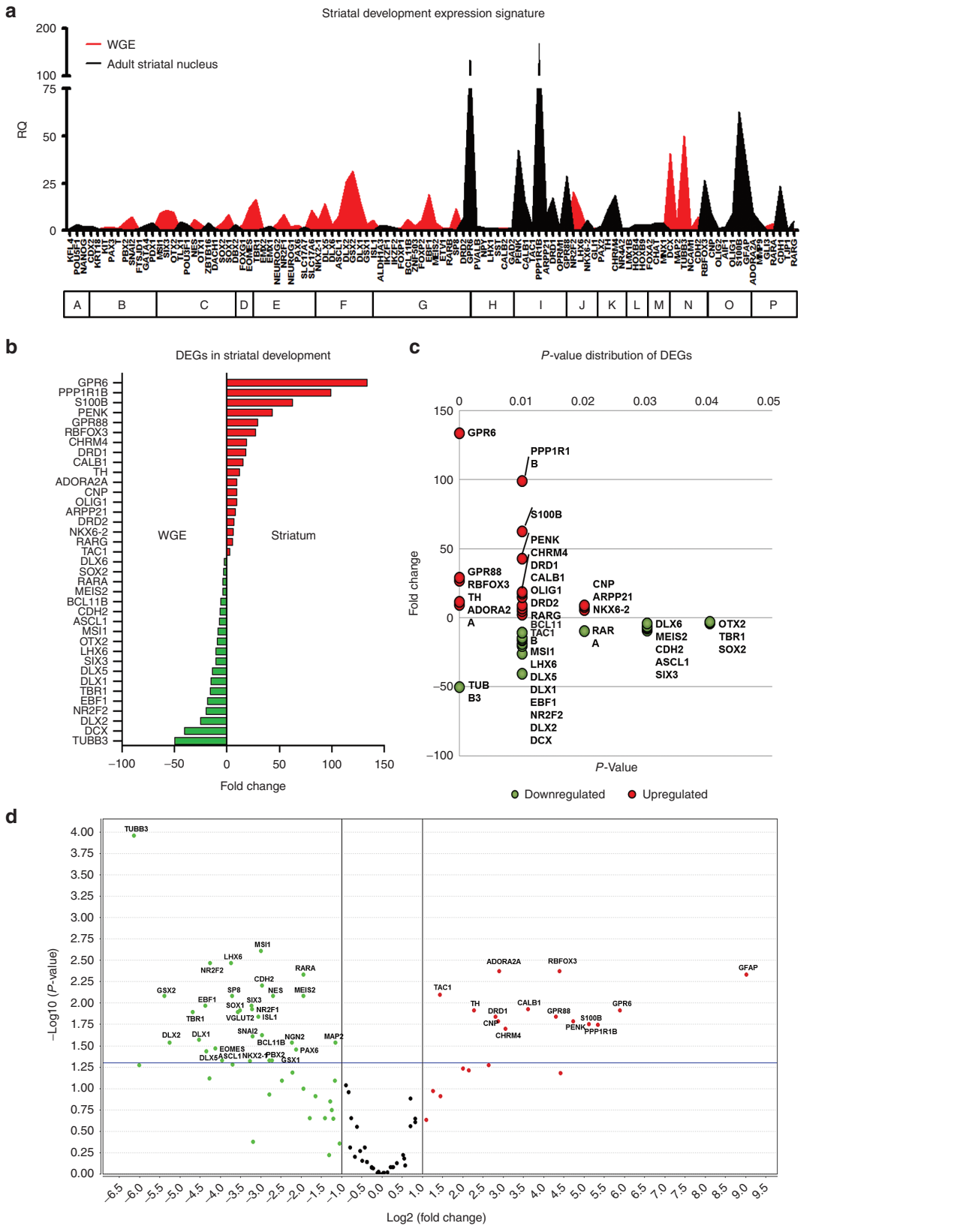
A quantitative gene expression tool for testing the efficiency of human pluripotent stem cell differentiation to MSNs

Stem cell telencephalic differentiation recapitulates the human neurodevelopmental process in the following three phases (Figure 6a): (i) Neuralization phase, where pluripotent stem cells are committed to neural cells via neuroectodermal induction; (ii) Patterning and specification phases, where rostro-caudal and dorso-ventral identities are established; (iii) Maturation, the final phase that generates neuronal subtypes and establish synaptic connectivity. We differentiated human PSC lines using a previously published protocol[11] and extracted RNA at 0, 4, 8, 16, 23, and 28 days *in vitro* (DIV) to quantify the expression levels of our panel of 106 selected genes using the same high-throughput quantitative PCR method that we used to evaluate gene expression during striatal and cortical development. To assess the efficiency of the differentiation protocol, we compared the expression data obtained from 0 to 28 DIV to that obtained from the 7–9 p.c.w. WGE and adult striatum samples with the latter data sets being used as developmental baselines. We also used the predictor genes that we identified above and their expression levels to distinguish subpallial from pallial derivatives in the differentiated PSC samples.

Unbiased hierarchical clustering analysis of the differentiated PSC, WGE, and adult striatum data sets produced two main clusters (Figure 6b) with the first cluster containing the samples from the pluripotent stage (0 DIV) until the rosette stage (8 DIV). These samples formed their own cluster partially due to the expression of pluripotency and germ layer marker genes that decrease from 0 DIV to 8 DIV and which is not significantly expressed beyond this time point (Supplementary Data 3A and B).

The second cluster contained all of the differentiated samples after 8 DIV as well as all of the WGE and adult striatal samples. Overall, the 16 DIV to 28 DIV samples clustered closer to the WGE group than the striatum group indicating that at these differentiation stages, the samples were more similar to the WGE. However, an interesting time-dependent clustering pattern was observed. The 16 DIV samples (DLX1-positive neural progenitors—Figure 6c) clustered closer to the WGE than did the 23 DIV samples (β-III tubulin-positive

Figure 3 Quantitative gene expression profiles of human fetal whole ganglionic eminence (WGE) and cortex between 7 and 9 p.c.w. (**a**) WGE and fetal cortex are compared for the expression of the 106 selected genes involved in CNS development. At this stage, fetal cortex (light blue peaks) shows expression of pallial markers (*EOMES, TBR1, EMX1* and *2*, *NEUROG2*) as well as low mRNA levels of VGLUT1 transporter (*SLC17A7*). Human WGE shows higher expression of early neural markers, especially *SIX3* and *OTX2*. As expected, subpallial and pallial markers are more highly expressed in WGE and fetal cortex respectively. WGE already shows increased expression of medium spiny neuron markers (*GAD2, TAC1, OPRM1*) compared to fetal cortex. (**b**) Statistically significant differentially expressed genes between WGE and fetal cortex are depicted. Gene expression levels are calculated as fold change relative to fetal cortex for upregulated genes (red histograms) and fold change relative to WGE for downregulated genes (green histograms). (**c**) Statistically significant genes identified in Figure 3b are plotted by *P* value (x-axis) and fold change (y-axis). *TAC1, DLX5,* and *DLX6* are the most significantly upregulated genes in the WGE compared to the fetal cortex with a Student's *t*-test and *P* value of <0.001. (**d**) Six genes are identified as the best predictors for WGE identity. Differential analysis of k-mean clustering analysis and Student's *t*-test, followed by one-way analysis of variance (ANOVA), shows *DLX1, DLX5, DLX6, EBF1, LHX6,* and *SIX3* are the main reliable identifier genes for WGE. One-way ANOVA followed by Neuman-Keuls multiple comparison test, $^{**}P < 0.01$; $^{***}P < 0.001$. WGE at 49 pcd has been used as internal control. Error bars = SEM. DEGs, differentially expressed genes; pcd, postconception days; p.c.w., postconception weeks; RQ, relative quantity.



neurons—Figure 6c). The 28 DIV samples were the most peripheral of this subcluster and the closest to the adult striatal sample cluster. Differentiated PSCs showed downregulation of *OTX2* after 8 DIV and of *SOX1* after 16 DIV (Supplementary Data 3C and D) indicative of the shutting down of the early neural gene expression program was turned down. The LGE/medial ganglionic eminence marker genes *GLI1*, *GSX1*, and *GSX2* were downregulated after 16 DIV indicating a switch toward a mature striatal fate (Supplementary Data 3E). This observation was supported by the upregulation of *GPR88* (Supplementary Data 3F), which we defined as one of the striatal

predictor genes (Figure 4b,c). Furthermore, an increase in *RBFOX3* (NeuN) expression that occurs between 16 and 23 DIV suggests the onset of neuronal maturation (Supplementary Data 3G).

We also performed immunocytochemistry to assess their progress through the various phases (Figure 6c). At the start of the differentiation protocol (0 DIV), all cells were pluripotent as shown by OCT4 staining, and by 4 DIV, the vast majority of the cells were neuralized as shown by the expression of the neuroectodermal marker Pax6. At 8 DIV, neuroepithelium-like structures known as rosettes[12] had formed with the cells comprising the rosettes expressing the rosette-specific marker PLZF and displaying a rosette-specific polarized subcellular distribution of ZO1. By 16 DIV, a significant proportion of the cells expressed DLX1 protein indicating that these cells had acquired a subpallial progenitor fate. At 23 DIV, practically all cells expressed β-III tubulin and showed a neuronal morphology. By 28 DIV, DARPP-32/CTIP2 double-positive cells with a more complex projection system were observed indicating that these cells were acquiring a MSN fate (Figure 6c).

DISCUSSION

In recent years, stem cell technology has been used successfully to establish *in vitro* model to study human brain biology[13] and has furthered our knowledge of neurodevelopment at the molecular and cellular levels. Furthermore, stem cell-derived tissues and differentiated cells have great potential in drug discovery and cell therapy for the treatment of a range of neurological disorders. However, the possibility of modeling specific human diseases *in vitro* or the generation of cells suitable for transplantation relies on the efficacy and efficiency of the differentiation protocols used. It is generally agreed that it is necessary to recapitulate *in vitro* the critical *in vivo* developmental stages of the cell type of interest to produce a robust differentiation protocol. Thus, it is necessary to develop tools to compare the gene expression profiles of cells derived by *in vitro* differentiation with the equivalent target tissues derived from normal human developmental samples. The differentiation of stem cells to MSNs is important to model system for many diseases that affect the basal ganglia, most notably Huntington's disease. With this in mind, the characterization of human striatal development is fundamental for understanding *in vivo* MSN specification for its subsequent recapitulation *in vitro*. Here we have determined the gene expression profile of a defined set of developmentally regulated genes in fetal and adult human striatal tissue by qRT-PCR. We demonstrate that the expression of this set of genes can be used to discriminate different developmental stages, and that the levels of expression of these genes are also relevant to the temporal maturation of the region. Subsequently we show that these tissue-specific gene expression signatures can be used to evaluate a protocol that differentiates PSCs to striatal neurons.

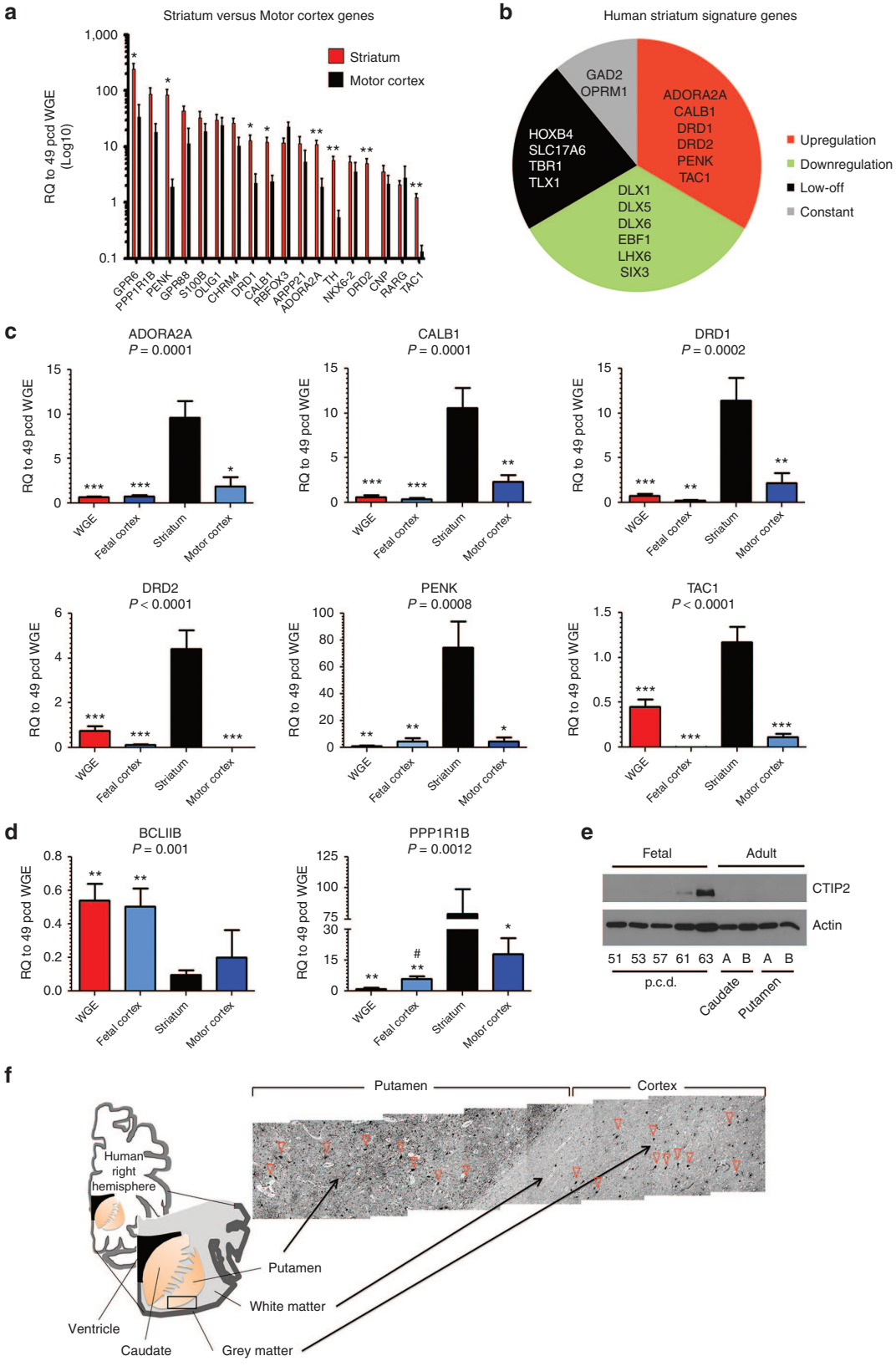
To establish a qRT-PCR array for use with OpenArray technology,[14] we selected 106 genes relevant to neural induction and telencephalic and subpallial lineage specification and differentiation. Use of this gene array was then first validated from the perspective of human development. We analyzed human fetal WGE and adult striatal samples to establish quantitative gene expression signatures of specified neural progenitors and mature neurons. Unbiased hierarchical clustering revealed that this set of genes did indeed distinguish between areas (WGE and cortical primordia) and developmental stages (fetal and adult). Our findings indicate that our gene set is sufficiently stringent to distinguish between pallial and subpallial derivatives. We used a qRT-PCR based analysis method; therefore, we could establish gene expression signatures based not only on whether or not certain genes were expressed but also on the size and statistical significance of changes in expression of individual genes in different tissues and at different developmental stages. Furthermore, as we simultaneously analyzed the quantitative expression of 106 genes, we can also consider the overall gene expression profile as part of the signature.

Relatively few transcriptomic studies of human forebrain and more specifically striatal development are currently available, which highlights the importance of comparing our results with previous published data. We compared our developmental quantitative gene expression data with microarray data present in the Human Brain Atlas (www.hbatlas.org)[2] and similar gene expression profiles were observed in both data sets (see *CTIP2* below as an example). The genes that we identified as highly specific markers for adult striatal neurons are also conserved during primate evolution as shown by transcriptomic analysis of adult human, chimpanzee, and macaque telencephalon.[15] This reinforces the accuracy of the adult striatal gene expression signature that we generated.

Multiple levels of analysis has allowed us to identify a gene expression signature, which consists of the following four classes of genes: (i) WGE genes, which show higher expression in the WGE than in the striatum (*DLX1, DLX5, DLX6, EBF1, LHX6,* and *SIX3*); (ii) striatal genes, which show higher expression in the adult striatum than in the WGE (*ADORA2A, CALB1, DRD1*, *DRD2, PENK,* and *TAC1*); (iii) constitutively expressing genes, that do not change their expression levels between WGE and adult striatum (*GAD2* and *OPRM1*); and (iv) low expression genes, that are expressed at much lower levels in the adult striatum compared to the other tissues tested (*TBR1, TLX1, SLC17A6,* and *HOXB4*).

The WGE gene cluster identifies genes that are highly expressed in the 7–9 p.c.w. WGE samples. Within this gene cluster, the most highly expressed genes were *SIX3* and *DLX* genes *DLX1*, *DLX5,* and *DLX6*, which are well characterized as transcription factor genes required for ventral telencephalic specification in mammals.[16–19] Interestingly *DLX2*, another *DLX* family member present and commonly used marker of LGE development, was not defined as a WGE-specific gene at this stage since its expression is only slightly higher in the WGE than in the fetal cortex. One plausible explanation could be that *DLX2* is expressed in the developing human brain at an earlier stage than *DLX1, DLX5,* and *DLX6* as has been observed in rodents.[18,19] Among the *DLX* genes, the highest expressed in the

Figure 4 Quantitative gene expression profile of caudate-putamen nuclei development. (**a**) Whole ganglionic eminence (WGE) and striatal expression profiles of the 106 gene set. The letter code below the graph identifies the subgroups of genes that define specific cell or tissue types (see Table 1). (**b**) Differentially expressed genes (DEGs) are identified by Student's *t*-test of adult striatum compared to WGE striatal upregulated genes (red histograms) are mainly medium spiny neuron-related genes together with mature neuronal and glial markers. Striatal downregulated genes (green histograms), specific for WGE identity, are mainly early ventral forebrain as well as early neuronal genes. The x-axis represents the fold change of upregulated and downregulated genes compared to human WGE and striatum, respectively. (**c**) Statistically significant genes identified in Figure 4b are plotted by *P* value (x-axis) and fold change (y-axis). *GPR6, GPR88, RBFOX3, TH,* and *ADORA2A* are the most significant upregulated genes in the striatum compared to the WGE, while *TUBB3* is the most significant downregulated gene. (**d**) Unbiased volcano plot analysis further identifies DEGs. *P* value boundary, indicated by a blue threshold line, is 0.05 (-log10 (*P* value) on the y-axis) and fold change boundary is 2 (log2(fold change)) on x-axis. Upregulated (red dots) and downregulated (green dots) significant genes are above the *P* value boundary line. Genes with similar expression levels (black dots) are plotted in the lower-central part of the volcano plot. RQ, relative quantity.



7–9 p.c.w. WGE samples are *DLX5* and *DLX6* which coincides with a previous human brain transcriptomic analysis.[2] This is consistent with the observation that *DLX5* and *DLX6* are highly expressed in mouse LGE around E11.5-E12.5.[18] Interestingly, although *GSX1* and *GSX2* appeared in the WGE cluster generated by the k-means clustering analysis, they did not show statistically significant difference in expression between the WGE and fetal cortex at 7–9 p.c.w. In mouse, these genes are expressed relatively early at around E10.5,[20,21] which suggests an early role in WGE development. A similar observation has been made during human development[2] and at

later stages only a thin subventricular zone layer can be observed at 18 gestational weeks.[22] Similar to *Dlx5* and *Dlx6*, mouse *Ebf1* expression starts at E11.5 where it regulates early striatal differentiation.[23] Analysis of human development by microarray also shows that *EBF1* expression peaks at around 9 p.c.w., and coincides with *DLX5* and *DLX6*.[2] Taken together, these results suggest that our OpenArray gene set is a powerful tool to determine an exact human developmental stage and indicate that human 7–9 p.c.w. WGE samples represent an early phase of striatal differentiation. This observation is strengthened by the fact that the WGE-specific gene set did not include *Ikaros* (*IKZF1*) and *Helios* (*IKZF2*), which we identified previously as markers of the later phase of striatal development.[24,25]

Interestingly, a previous microarray analysis of human fetal striatum from 18 to 23 p.c.w.[3] demonstrated the expression of genes that we identified as reliable specific markers of either WGE or adult striatum, indicating the developmental transition from WGE to striatum at this developmental stage. Our results identified *ADORA2A* as an adult striatal marker, which agrees with previous work that identified it as a hub gene in human caudate nucleus development.[15] We also identified *CALB1, PENK,* and the G protein-coupled receptor genes *GPR6* and *GPR88* as adult striatum specific genes. This demonstrates an evolutionary conservation of striatal specific genes as transcriptomic analysis of mouse striato-pallidal neurons also showed specific expression of *Adora2a, Bcl11b (Ctip2), Calb1, Gpr6, Gpr88,* and *Penk*.[26,27] Interestingly, our results did not identify *PPP1R1B* (*DARPP-32*) as a specific striatal marker since it is also expressed at a relatively high level in both fetal and adult cortex. Although these results seem at odds with previously published work,[28] this discrepancy can be explained by an increase in expression variability due to differences in *PPP1R1B* mRNA levels in our human specimens. *CTIP2* is another well-expressed gene in rodent MSN,[29] but here we have observed low *CTIP2* mRNA expression and negligible CTIP2 protein expression in adult striatal tissue, which also paralleled data in the Human Brain Atlas.[30] We hypothesize that CTIP2, when colocalizing with DARPP-32, may in fact be a marker of MSNs at the human fetal developmental stage, but not in the adult human striatum. This would explain its classification as a MSN marker in other studies where only human fetal brain tissue was analyzed.[4]

The striatal interneuron markers *TH* and *CHAT*[31,32] were also identified as striatal markers in this analysis. However, their expression could also include signal from neurons projecting from the substantia nigra and cortex with their projections, being known to contain mRNA of these genes.[33]

While protocols to generate MSNs from PSCs have been published previously, they generate MSNs with varying degrees of success.[4–6,11,34,35] Having established gene expression signatures from both WGE and adult striatum samples our aim was to use the OpenArray as a tool to evaluate the degree of striatal differentiation that was achieved using PSC telencephalic differentiation protocol.[11] OpenArray gene expression profiling to evaluate PSC differentiation protocols differs from the more commonly used analytical approaches where immunocytochemistry and qRT-PCR are performed for much smaller groups of selected markers.[4,5,36] More advanced transcriptomic analyses using techniques such as microarray or RNAseq are also viable approaches for monitoring the progress of PSC cultures through differentiation protocols.[4,36] However, due to the large amounts of data that are generated, data analysis can be time-consuming. Together with the higher cost, it is not feasible to apply these approaches (*i.e.*, RNA-seq) for iterative analysis of cultures to evaluate changes made during protocol development, nor to check the reproducibility of protocols within a research group (intrareproducibility) or between research groups (inter-reproducibility).

Our analysis of cultures at different time points during the human PSCs differentiation protocol revealed their developmental progression and identified the following three distinct developmental phases: (i) A presubpallial specification phase (0–8 DIV); (ii) A subpallial specification phase (8–16 DIV); (iii) A postsubpallial phase where the cells resemble WGE that is progressing toward an adult striatal fate (16–28 DIV). Importantly, the data demonstrated that the differentiation protocol we used[11] is a reliable telencephalic differentiation protocol that by 16 DIV produces ventral forebrain-subpallial progenitors with a gene expression profile similar to 7–9 p.c.w. human WGE.

Comparison of the final time point examined (28 DIV) with the adult striatal gene expression signature revealed that at 28 DIV the neurons were not yet approaching a fully differentiated mature phenotype. This suggests that further protocol optimization may be required. Once the neurons arrive at the final stages of differentiation and maturation, it will then be possible to compare them with the adult striatal gene expression profile, with a focus on relative changes in expression of the 18 adult striatal signature genes.

In this study, we show that quantitative gene expression profiles are useful for the comparison of the various stages of neurodevelopment, especially since the levels of specific striatal transcripts change between embryonic and adult stages. Our methods and findings described here enhance our understanding of human brain development, and provide a highly reliable, fast and low-cost approach for evaluating PSC-derived MSNs, which in turn will facilitate the development of new therapeutic and may give us a deeper understanding of the cellular and molecular mechanisms underlying diseases in which MSNs are dysfunctional or degenerate, such as Huntington's disease. Moreover, the approach described here could be used to establish gene expression signatures for many different *in vivo* tissues with these signatures then being available to

Figure 5 A specific set of genes distinguishes human adult striatum from whole ganglionic eminence (WGE). (**a**) Genes that are specifically upregulated in adult striatum compared to motor cortex (two-tailed Student's *t*-test, $*P < 0.05$, $**P < 0.01$). (**b**) The specific behavior of the genes that best identify adult human striatum. The upregulation, downregulation, and low-off subsets are defined as such relative to expression in cortical structures and WGE. The constant gene subset is defined as genes that show similar expression in both the WGE and adult striatum. (**c**) At the gene expression level, *ADORA2A, CALB1, DRD1, DRD2, PENK,* and *TAC1* are the best predictors for medium spiny neuron specification compared to cortical structures and WGE as shown by *P* values of one-way analysis of variance (ANOVA) analysis. WGE of 49 pcd has been used as internal control. (**d**) BCL11B (CTIP2) and PPP1R1B (DARPP-32) are the two main protein markers for MSN identification during striatal specification. At the fetal stages tested, BCL11B gene expression is equally high in both fetal cortex and WGE, while in adult striatal tissue expression is decreased due to striatal neuronal maturation. *PPP1R1B* expression displays the opposite trend and shows lower expression levels in fetal tissues as well as in adult motor cortex, suggesting a specific increase of expression during the development and maturation of human striatum. One-way ANOVA followed by Newman-Keuls multiple comparison test is applied, $*P < 0.05$; $**P < 0.01$ compared to striatum; $***P < 0.001$ compared to striatum. $\#P < 0.05$ compared to fetal cortex. Error bars = SEM. (**e**) Western blot analysis shows that CTIP2 protein is detectable in human fetal samples at 61 and 63 postconception days (pcd) with increasing intensity but is not detected in adult samples (Caudate and Putamen). Actin is used as the loading control. (**f**) Immunohistochemistry analysis in coronal brain sections of human adult ventral region of putamen and the adjacent insular cortex reveals DARPP-32-positive cells either in striatal and cortical tissue (white and gray matter). Red outlined arrow-heads highlight some of the frequent DARPP-32-expressing cells across the brain parenchyma. MSN, medium spiny neuron; RQ, relative quantity.

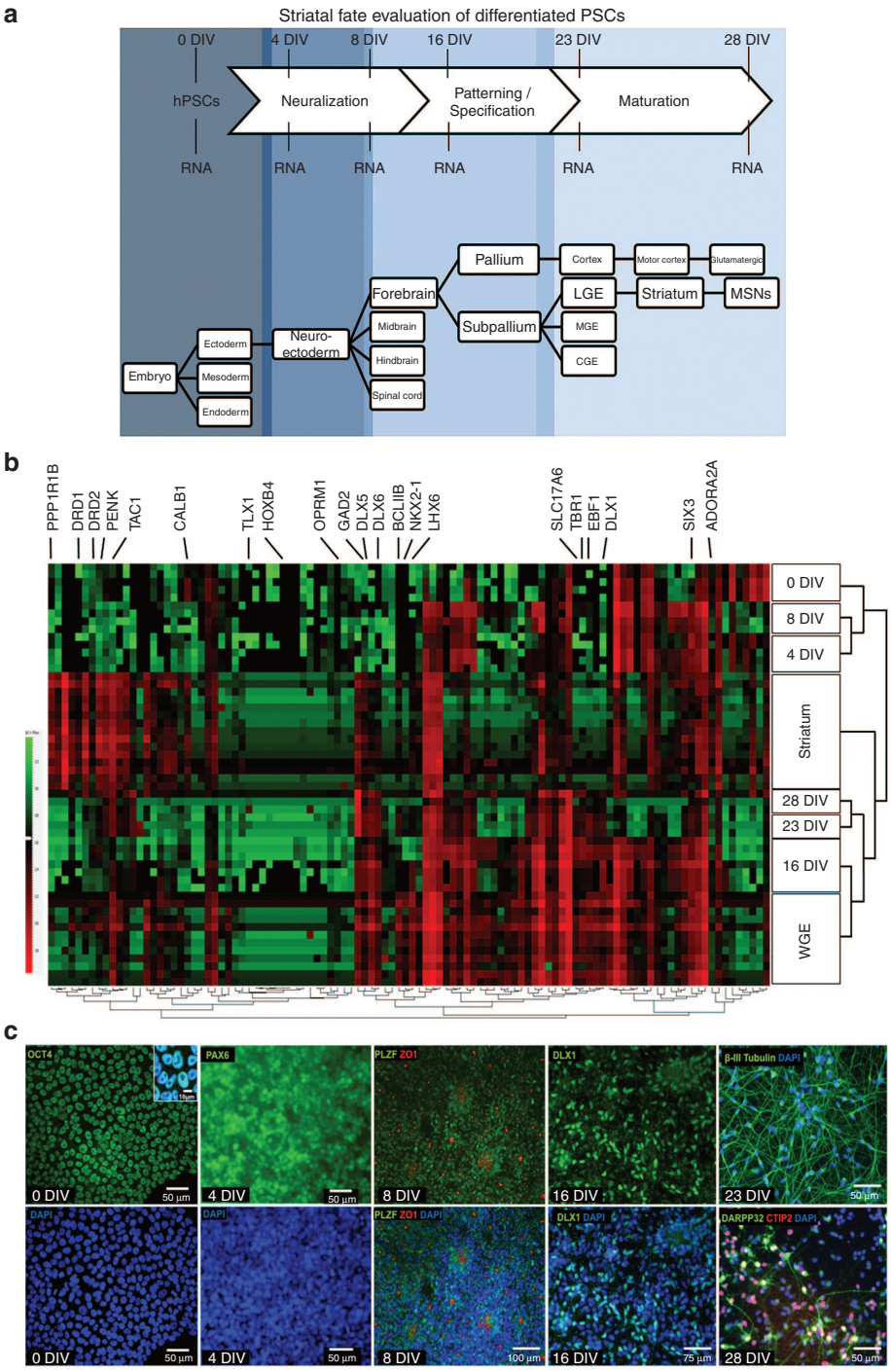


Figure 6 The striatal development gene expression signature is applied to evaluate the efficiency of human pluripotent stem cell (hPSC) differentiation toward a striatal fate. (**a**) Differentiation phases from pluripotent stem cells (PSCs) to medium spiny neurons (MSNs). Stem cell differentiation protocols translate *in vivo* development into an *in vitro* system. PSC-derived neurons are obtained via a three-step protocol. The first step is neuroectodermal induction (neuralization) that generates neural progenitors. The second step confers rostro-caudal and dorso-ventral patterning occurs and the final step involves the neurogenesis and differentiation of neuronal subtypes. This is a representation of how this three-step process mimics subpallial and then striatal development to generate MSNs. (**b**) Unbiased hierarchical clustering of differentiated hPSCs with human whole ganglionic eminence (WGE) and adult striatum. We isolated RNA from differentiating human PSCs at six time points (shown in panel **a**) to analyze the progress of the differentiation protocol by high-throughput quantitative real-time PCR. We compared the gene expression profile of differentiated hPSCs (H9 and hiPSC) to human WGE and adult striatal tissue in order to establish the spatio-temporal gene expression pattern of iPSC-derived neural progenitors and neurons. Striatal signature genes with *PPP1R1B* and *BCL11B* are indicated in the heat map. (**c**) Representative immunofluorescence images of hESCs (H9) and hiPSCs (CS83iCTR33n1) during differentiation from pluripotency to striatal neurons. H9 cells were differentiated until 16 DIV, while hiPSCs were differentiated to 28 DIV. At 4 DIV, the majority of cells had a neuroectodermal identity as shown by the PAX6 staining. At 8DIV, a neuroepithelial morphology was observed with the cells organizing in rosettes, expressing nuclear marker PLZF and displaying polarized ZO1 distribution. At 16 DIV, subpallial progenitors expressed DLX1 and give rise to β-III tubulin positive neurons at 23 DIV. At 28 DIV, the presence of DARPP-32/CTIP2 (MSN markers at the protein level) double-positive neurons is observed. CGE, caudal ganglionic eminence; DIV, days *in vitro*; hESC, human embryonic stem cell; hiPSC, human induced pluripotent stem cell; LGE, lateral ganglionic eminence; MGE, medial ganglionic eminence.

benchmark differentiation protocols that aim to produce these tissues *in vitro*.

MATERIALS AND METHODS

Postmortem human brain tissue

Adult motor cortex, caudate, and putamen samples were obtained from neurologically healthy brain donor at the Neurological Tissue Bank of the Biobank-Hospital Clinic—Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS; Barcelona, Spain; URL: www.clinicbiobanc.org), following the guidelines and approval of the local ethics committee. Human fetal tissue (CRL of 22–54 mm) ranging in age from 7–9 p.c.w was collected by donation of the products of the elective termination of pregnancy, within the MRC- and Welsh government-sponsored South Wales Initiative for Fetal Tissue Transplantation ("SWIFT") program, with full ethical approval (02/4446) For mRNA studies, cortex and striatal primordia were processed as previously published.[37] Patients' gender and ages are provided in Supplementary Data 1A.

RNA isolation and retrotranscription

Total RNA was isolated using TRI Reagent (T9424, Sigma-Aldrich, Madrid, Spain) or Trizol (15596, Life Technologies, Barcelona, Spain) following the manufacturers' protocol. Ten microliters of total RNA at a concentration of 200 ng/µl (2 µg in total) for each sample were reverse-transcribed with random primers using the High-Capacity RNA-to-cDNA Kit (4387406, Life Technologies, Barcelona, Spain). Ten microliters of retro-transcription cocktail (2 µl of 10× retrotranscription (RT) buffer, 2 µl of Random primers, 1 µl of dNTP mix; 1 µl MultiScribe reverse transcriptase) were added to each sample (20 µl total volume). After gentle mixing, the samples were incubated for 10 minutes at room temperature followed by 2 hours at 37 °C, 10 minutes on ice, and 10 minutes at 75 °C.

Openarray

Customized Openarrays (Life Technologies, Barcelona, Spain) containing 112 TaqMan probes (see Supplementary Data File 1), that specifically detect all isoforms of each of the 106 genes that were selected from the literature, as well as six housekeeping genes (*18S, B2M, HPRT1, HSP90AB1, RPL13A, UBC*) that were used as reference genes were produced and validated. cDNAs were loaded onto the custom OpenArrays and run as recommended by the manufacturer on the QuantStudio 12K Flex Real-Time PCR system by Servei Veterinari de Genètica Molecular at Faculty of Veterinary in Universitat Autònoma de Barcelona (Spain).

Data analysis

Human WGE samples from 49 to 64 ultrasound pcd (postconceptional day) were consider as the WGE biological group in order to acquire more statistical power, with the caveat of losing biological information during this fetal development period. Relative gene expression values were calculated using Expression Suite Software 1.03 (Life Technologies, Barcelona, Spain). RQ minimum and maximum values (error bars) are calculated with a confidence level of 95%, using Benjamini-Hochberg false discovery rate to adjust *P* values. Maximum allowed C_t included in calculations is 30 and C_q confidence > 0.8. Multivariate Student's *t*-test or one-way ANOVA were applied and values of $P < 0.05$ were considered statistically significant. Error bars are presented in all graphs as standard error of the mean (SEM). Gene expression profile data are represented in graphics as relative quantity or fold change of one biological group normalized to another. Volcano plot analysis comparing WGE and striatal samples was performed with Data Assist v3.01 software (Applied Biosystem) using the same parameters described above for the Expression Suite software.

Heat maps generation and k-means clustering

Double ΔC_t values ($2^{-\Delta\Delta C_t}$) were log2 transformed and normalized to ±3 SD for representation on the heatmap. The complete hierarchical cluster algorithm and k-means clustering were applied using the Genesis software from the Institute for Genomics and Bioinformatics (University of Graz). Global view map type for unbiased hierarchical clustering for all 106 genes to compare human tissue with hiPSCs in differentiation was generated with Data Assist v3.01 software (Applied Biosystem) using Pearson's correlation for distance measure and complete linkage as clustering method.

Protein extraction and western blot

Total protein extract was isolated from human tissue using TRI Reagent (T9424, Sigma-Aldrich, Madrid, Spain) according to the manufacturer's protocol. Total protein extracts were denatured using 2.5 mmol/l dithiothreitol at 100 °C for 5 minutes and then 40 µg of each denatured samples was subjected to 12% SDS-PAGE and transferred to a nitrocellulose membrane (IPVH00010, Millipore, Barcelona, Spain) for 90 minutes at 1 mA/cm^2 as described previously.[38] Membranes were incubated with polyclonal rabbit anti-Ctip2 (overnight at +4 °C, dilution of 1 µg/ml; Abcam Ab28448) and anti-Actin (20 minutes at RT, dilution of 1:20,000, MP Biochemicals), diluted in immunoblot buffer (Tris-buffered saline (TBS) containing 0.05% Tween-20 and 5% no-fat dry milk).

Immunocytochemistry

Cultured cells were fixed with 4% paraformaldehyde in phosphate-buffered saline (PBS) for 20 minutes at room temperaure, and then washed with PBS. Cells were incubated overnight at 4 °C with 7% normal goat serum (S-1000, VectorLabs, Peterborough, UK) in PBS containing 0.3% Triton ×100, 1% Bovine serum albumin (BSA), 0.3% sodium azide and primary antibodies: mouse anti-OCT4 (1:100, sc-5279, Santa Cruz Biotechnology, Heidelberg, Germany), mouse anti-PAX6 (1:200; Hybridoma Bank, Iowa City, Iowa), mouse anti-PLZF (1:100, OP128, Calbiochem, Madrid, Spain), rabbit anti-ZO1 (1:100, 40–2200, Invitrogen, Barcelona, Spain), rabbit anti-DLX (1:200, PA5967, CHDI, Princeton, NJ), anti-rabbit DARPP-32 (1:100, sc-11365, Santa Cruz Biotechnology, Heidelberg, Germany), anti-rat CTIP2 (1:300, Abcam ab18465), anti-rabbit β-III tubulin (1:1000, T2200, Sigma-Aldrich, Madrid, Spain). After rinsing in PBS, cells were incubated for 1 hour at RT with secondary antibodies: goat anti-mouse Alexa 546 (1:1000, A-11018, Molecular Probes, Madrid, Spain), goat anti-rabbit Alexa 546 (1:1000, A-11010, Molecular Probes, Madrid, Spain), Alexa 488 (1:1000, A-11070, Molecular Probes, Madrid, Spain) or goat anti-rat Alexa 488 (1:500, A-11006, Molecular Probes, Madrid, Spain). Microscopy images were obtained using a LEICA AF6000 microscope with 10x and 20× objectives.

Immunohistochemistry

Five-micrometer thick sections were obtained from 4% formalin-fixed and paraffin-embedded tissue blocks from the caudate nucleus of a neurologically healthy brain donor at the Neurological Tissue Bank of the Biobank-Hospital Clinic-IDIBAPS, after obtaining informed consent for the use of brain tissue for diagnostic and research purposes. Immunohistochemistry was performed on an automated immunostainer (Dako Autostainer Plus, Barcelona, Spain) using anti-rabbit DARPP-32 (1:100, sc-11365, Santa Cruz Biotechnology, Heidelberg, Germany). Immunoreaction was visualized using the DAKO Envision system. For antigen retrieval, paraffin sections were dewaxed and boiled in a microwave for 20 minutes in citrate buffer at pH 6.

Striatal differentiation of human pluripotent stem cells

H9 human embryonic stem cells,[39] (WA09, WiCell, Madison, WI) Feeder Independent–TeSR1 Medium, and CS83iCTR33n1 hiPSCs 7, obtained from CHDI foundation biorepository at Coriell Institute, were differentiated until 28 DIV following the telencephalic differentiation protocol previously described.[11] All cells were regularly tested for mycoplasma contamination and their karyotype was verified routinely.

ACKNOWLEDGMENTS

We would like to thank Maria Calvo and Anna Bosch from the Confocal Microscopy Unit at the Scientific and Technological Centers of the University of Barcelona for their support. We also thank Anna Mercadé and Olga Francino form Servei Veterinari de Genètica Molecular (Univeritat Autonoma de Barcelona) for their assistance. We are indebted to the Neurological Tissue Bank of the Biobanc-Hospital Clinic-IDIBAPS for technical assistance and sample procurement. We also thank Ellen Gelpi for helpful suggestions. Thanks to Victoria H. Roberton for her help in collection of fetal tissues and Mercè Masana for critical review of the manuscript. This study was supported by grants from the Ministerio de Economia y Competitividad (SAF2011-29507 to J.A.; SAF2012-37417 to J.M.C.) and from the ISCIII-Subdirección General de Evaluación and European Regional Development Fund (ERDF) (RETICS to J.M.C. (RD12/0019/0002; Red de Terapia Celular)), Spain; and CHDI Foundation (A-4528 to N.D.A. and A-7332 to J.M.C.), USA. iPSCs were obtained and characterized by the iPSCs-Core from CEDARS-Sinai (Los Angeles, CA, USA) in the context of the HD-iPSC Consortium supported by the NINDS and CHDI Foundation, USA. M.S.,

P.S., G.B., J.C., S.Y., C.N.S., P.J.K., J.A., R.C.S., N.D.A., and J.M.C. are members of the HD-iPSC Consortium. M.S. designed the study, performed the majority of the experiments, and drafted the manuscript. G.G.-D.B. and J.A. performed hierarchical clustering and k-means analyses. P.S. participated in stem cell differentiation experiments and drafted the manuscript. S.Y., N.D.A., P.J.K., and G.B. participated in stem cell differentiation experiments. J.C. and P.B.M. participated in protein analysis. N.-N.V., C.M.K., and A.E.R. performed human fetal tissue dissection and recovery. C.N.S. characterized the human iPSCs. J.A. and R.C.S. participated in the study design. J.M.C. coordinated the study and drafted the manuscript. All authors critically revised and approved the final manuscript. The authors declare no financial or nonfinancial competing interests.

REFERENCES

1. Yamanaka, S and Blau, HM (2010). Nuclear reprogramming to a pluripotent state by three approaches. *Nature* **465**: 704–712.
2. Kang, HJ, Kawasawa, YI, Cheng, F, Zhu, Y, Xu, X, Li, M *et al.* (2011). Spatio-temporal transcriptome of the human brain. *Nature* **478**: 483–489.
3. Johnson, MB, Kawasawa, YI, Mason, CE, Krsnik, Z, Coppola, G, Bogdanović, D *et al.* (2009). Functional and evolutionary insights into human brain development through global transcriptome analysis. *Neuron* **62**: 494–509.
4. Delli Carri, A, Onorati, M, Lelos, MJ, Castiglioni, V, Faedo, A, Menon, R *et al.* (2013). Developmentally coordinated extrinsic signals drive human pluripotent stem cell differentiation toward authentic DARPP-32+ medium-sized spiny neurons. *Development* **140**: 301–312.
5. Aubry, L, Bugi, A, Lefort, N, Rousseau, F, Peschanski, M and Perrier, AL (2008). Striatal progenitors derived from human ES cells mature into DARPP32 neurons *in vitro* and in quinolinic acid-lesioned rats. *Proc Natl Acad Sci USA* **105**: 16707–16712.
6. Ma, L, Hu, B, Liu, Y, Vermilyea, SC, Liu, H, Gao, L, *et al.* (2012). Human embryonic stem cell-derived GABA neurons correct locomotion deficits in quinolinic acid-lesioned mice. *Cell Stem Cell* 10: 455–464.
7. The HD iPSC Consortium (2012). Induced pluripotent stem cells from patients with huntington's disease show cag-repeat-expansion-associated phenotypes. *Cell Stem Cell* **11**: 264–278.
8. Denham, M and Dottori, M (2009). Signals involved in neural differentiation of human embryonic stem cells. *Neurosignals* **17**: 234–241.
9. Evans, AE, Kelly, CM, Precious, SV and Rosser, AE (2012). Molecular regulation of striatal development: a review. *Anat Res Int* **2012**: 106529.
10. Ouimet, CC, Miller, PE, Hemmings, HC Jr, Walaas, SI and Greengard, P (1984). DARPP-32, a dopamine- and adenosine 3':5'-monophosphate-regulated phosphoprotein enriched in dopamine-innervated brain regions. III. Immunocytochemical localization. *J Neurosci* **4**: 111–124.
11. Martín-Flores, N, Romaní-Aumedes, J, Rué, L, Canal, M, Sanders, P, Straccia, M *et al.* (2015). RTP801 Is Involved in Mutant Huntingtin-Induced Cell Death. *Mol Neurobiol* (e-pub ahead of print).
12. Elkabetz, Y, Panagiotakos, G, Al Shamy, G, Socci, ND, Tabar, V and Studer, L (2008). Human ES cell-derived neural rosettes reveal a functionally distinct early neural stem cell stage. *Genes Dev* **22**: 152–165.
13. Lancaster, MA, Renner, M, Martin, CA, Wenzel, D, Bicknell, LS, Hurles, ME *et al.* (2013). Cerebral organoids model human brain development and microcephaly. *Nature* **501**: 373–379.
14. Patel, SN, Wu, Y, Bao, Y, Mancebo, R, Au-Young, J and Grigorenko, E (2013). TaqMan® OpenArray® high-throughput transcriptional analysis of human embryonic and induced pluripotent stem cells. *Methods Mol Biol* **997**: 191–201.
15. Konopka, G, Friedrich, T, Davis-Turak, J, Winden, K, Oldham, MC, Gao, F *et al.* (2012). Human-specific transcriptional networks in the brain. *Neuron* **75**: 601–617.
16. Carlin, D, Sepich, D, Grover, VK, Cooper, MK, Solnica-Krezel, L and Inbal, A (2012). Six3 cooperates with Hedgehog signaling to specify ventral telencephalon by promoting early expression of Foxg1a and repressing Wnt signaling. *Development* **139**: 2614–2624.
17. Lavado, A, Lagutin, OV and Oliver, G (2008). Six3 inactivation causes progressive caudalization and aberrant patterning of the mammalian diencephalon. *Development* **135**: 441–450.
18. Eisenstat, DD, Liu, JK, Mione, M, Zhong, W, Yu, G, Anderson, SA *et al.* (1999). DLX-1, DLX-2, and DLX-5 expression define distinct stages of basal forebrain differentiation. *J Comp Neurol* **414**: 217–237.
19. Liu, JK, Ghattas, I, Liu, S, Chen, S and Rubenstein, JL (1997). Dlx genes encode DNA-binding proteins that are expressed in an overlapping and sequential pattern during basal ganglia differentiation. *Dev Dyn* **210**: 498–512.
20. Toresson, H and Campbell, K (2001). A role for Gsh1 in the developing striatum and olfactory bulb of Gsh2 mutant mice. *Development* **128**: 4769–4780.
21. Hsieh-Li, HM, Witte, DP, Szucsik, JC, Weinstein, M, Li, H and Potter, SS (1995). Gsh-2, a murine homeobox gene expressed in the developing brain. *Mech Dev* **50**: 177–186.
22. Ma, T, Wang, C, Wang, L, Zhou, X, Tian, M, Zhang, Q *et al.* (2013). Subcortical origins of human and monkey neocortical interneurons. *Nat Neurosci* **16**: 1588–1597.
23. Garel, S, Marín, F, Grosschedl, R and Charnay, P (1999). Ebf1 controls early cell differentiation in the embryonic striatum. *Development* **126**: 5285–5294.
24. Martín-Ibáñez, R, Crespo, E, Urbán, N, Sergent-Tanguy, S, Herranz, C, Jaumot, M *et al.* (2010). Ikaros-1 couples cell cycle arrest of late striatal precursors with neurogenesis of enkephalinergic neurons. *J Comp Neurol* **518**: 329–351.
25. Martín-Ibáñez, R, Crespo, E, Esgleas, M, Urban, N, Wang, B, Waclaw, R *et al.* (2012). Helios transcription factor expression depends on Gsx2 and Dlx1&2 function in developing striatal matrix neurons. *Stem Cells Dev* **21**: 2239–2251.
26. Doyle, JP, Dougherty, JD, Heiman, M, Schmidt, EF, Stevens, TR, Ma, G *et al.* (2008). Application of a translational profiling approach for the comparative analysis of CNS cell types. *Cell* **135**: 749–762.
27. Lobo, MK, Karsten, SL, Gray, M, Geschwind, DH and Yang, XW (2006). FACS-array profiling of striatal projection neuron subtypes in juvenile and adult mouse brains. *Nat Neurosci* **9**: 443–452.
28. Onorati, M, Castiglioni, V, Biasci, D, Cesana, E, Menon, R, Vuono, R *et al.* (2014). Molecular and functional definition of the developing human striatum. *Nat Neurosci* **17**: 1804–1815.
29. Arlotta, P, Molyneaux, BJ, Jabaudon, D, Yoshida, Y and Macklis, JD (2008). Ctip2 controls the differentiation of medium spiny neurons and the establishment of the cellular architecture of the striatum. *J Neurosci* **28**: 622–632.
30. Sunkin, SM, Ng, L, Lau, C, Dolbeare, T, Gilbert, TL, Thompson, CL *et al.* (2013). Allen Brain Atlas: an integrated spatio-temporal portal for exploring the central nervous system. *Nucleic Acids Res* **41**(Database issue): D996–D1008.
31. Unal, B, Ibáñez-Sandoval, O, Shah, F, Abercrombie, ED and Tepper, JM (2011). Distribution of tyrosine hydroxylase-expressing interneurons with respect to anatomical organization of the neostriatum. *Front Syst Neurosci* **5**: 41.
32. Lapper, SR and Bolam, JP (1992). Input from the frontal cortex and the parafascicular nucleus to cholinergic interneurons in the dorsal striatum of the rat. *Neuroscience* **51**: 533–545.
33. Kuhl, D and Skehel, P (1998). Dendritic localization of mRNAs. *Curr Opin Neurobiol* **8**: 600–606.
34. El-Akabawy, G, Medina, LM, Jeffries, A, Price, J and Modo, M (2011). Purmorphamine increases DARPP-32 differentiation in human striatal neural stem cells through the Hedgehog pathway. *Stem Cells Dev* **20**: 1873–1887.
35. Kirik, D, Georgievska, B and Björklund, A (2004). Localized striatal delivery of GDNF as a treatment for Parkinson disease. *Nat Neurosci* **7**: 105–110.
36. Danjo, T, Eiraku, M, Muguruma, K, Watanabe, K, Kawada, M, Yanagawa, Y *et al.* (2011). Subregional specification of embryonic stem cell-derived ventral telencephalic tissues by timed and combinatory treatment with extrinsic signals. *J Neurosci* **31**: 1919–1933.
37. Kelly, CM, Precious, SV, Torres, EM, Harrison, AW, Williams, D, Scherf, C *et al.* (2011). Medical terminations of pregnancy: a viable source of tissue for cell replacement therapy for neurodegenerative disorders. *Cell Transplant* **20**: 503–513.
38. Straccia, M, Dentesano, G, Valente, T, Pulido-Salgado, M, Solà, C and Saura, J (2013). CCAAT/enhancer binding protein β regulates prostaglandin E synthase expression and prostaglandin E2 production in activated microglial cells. *Glia* **61**: 1607–1619.
39. Thomson, JA, Itskovitz-Eldor, J, Shapiro, SS, Waknitz, MA, Swiergiel, JJ, Marshall, VS *et al.* (1998). Embryonic stem cell lines derived from human blastocysts. *Science* **282**: 1145–1147.

Supplementary Information accompanies this paper on the *Molecular Therapy—Methods & Clinical Development* website (http://www.nature.com/mtm)