Creatio | Production and validation center of advanced therapies
UNIVERSITAT DE BARCELONA

In vitro study of neurodevelopment in Huntington's disease

Phil Sanders, Waseem Abbas, Anna Esteve-Codina, Gustavo Rodriguez-Esteban, Georgina Bombau, Mireia Galofré, Andrea Honrubia, Holger Heyn, Petia Radeva, & Josep M. Canals

UNIVERSITAT DE BARCELONA

Abstract

Huntington's disease (HD) is a neurodegenerative disorder that primarily affects striatal medium spiny neurons (MSNs). Recent evidence indicates that there is a neurodevelopmental component to HD where MSN specification, maturation and homeostasis may be affected[1-5]. However, the dysregulated cellular mechanisms underlying these developmental impairments remain to be established. To enhance understanding of how neurodevelopment is affected in HD, further study of striatal development in a healthy context is also required as it is a relatively understudied brain region.

Our aim is to study human striatal development in both a healthy and HD context in parallel. To achieve this, we use an *in vitro* model of striatal development where control and HD human pluripotent stem cells are differentiated towards a MSN fate, and analysed at various stages during the differentiation process using a range of transcriptomic and functional assay methods.

By integrating the data that is obtained we will identify striatal cell types and the cellular pathways and processes required for their specification, in addition to obtaining insight into the developmental dynamics of these cell types. We will also gain insight into how the development of these cell types is altered in HD and potentially identify targets whose modulation will restore MSN specification, homeostasis and function to a healthy state.

1. van der Plas *et al* (2019). *Neurology.* **93**, e1021-e1030.
2. Barnat *et al* (2020). *Science.* **369**, 787-793.
3. The HD iPSC Consortium (2017). *Nature Neuroscience*. **20**, 648-660.
4. Marder & Mehler (2012). *Neurology.* **79**, 621-622.
5. Mehler & Gokhan (2000). *Trends Neurosci.* **23**, 599-605.

Methods

Cell lines: Genea19 (GEN-019 – control) and Genea20 (GEN-020 – HD, 48 CAG repeats) human embryonic stem cell lines.

***In vitro* differentiation:** GEN-019 and GEN-020 were differentiated using our forebrain neuron *in vitro* differentiation protocol[6-8].

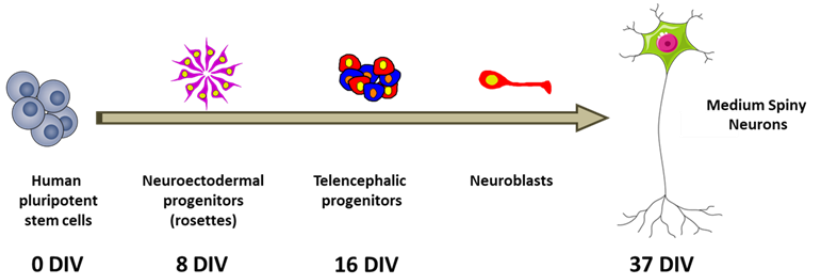
Bulk RNA-seq: Gene and isoform differential expression analysis were performed using DESeq2/1.18 and RSEM respectively. Functional enrichment analysis was performed using g:Profiler[9].

Single cell RNA-seq: Performed using the Smart-Seq2 protocol[10]. Data analysis and clustering analysis were performed using Seurat[11] and UMAP[12] respectively.

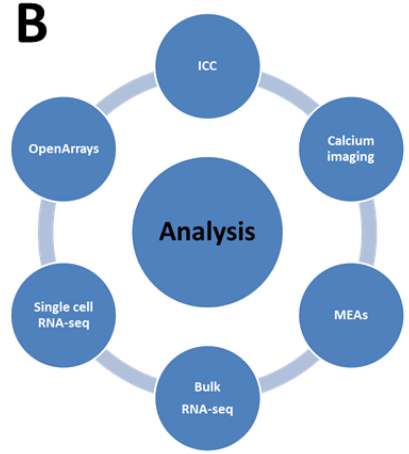
Machine learning: Analyses were performed using custom developed scripts powered by Keras, Scanpy, Scikit-learn and Scipy Python libraries.

6. Straccia *et al* (2015). *Mol. Ther. Methods Clin. Dev.* **2**, 15030.
7. Telezhkin *et al* (2015). *Am. J. Physiol. Cell Physiol.* **310**, C520-C541.
8. Comella-Bolla *et al* (2020). Mol. Neurobiol. **57**, 2766-2798.
9. Raudvere *et al* (2019). Nucleic Acids Res. **47**(W1), W191-W198.
10. Picelli *et al* (2014). *Nature Protocols.* **9**, 171-181.
11. Stuart *et al* (2019). *Cell* **177**, 1888-1902.
12. McInnes *et al* (2018). arxiv.org/abs/1802.03426

A



B

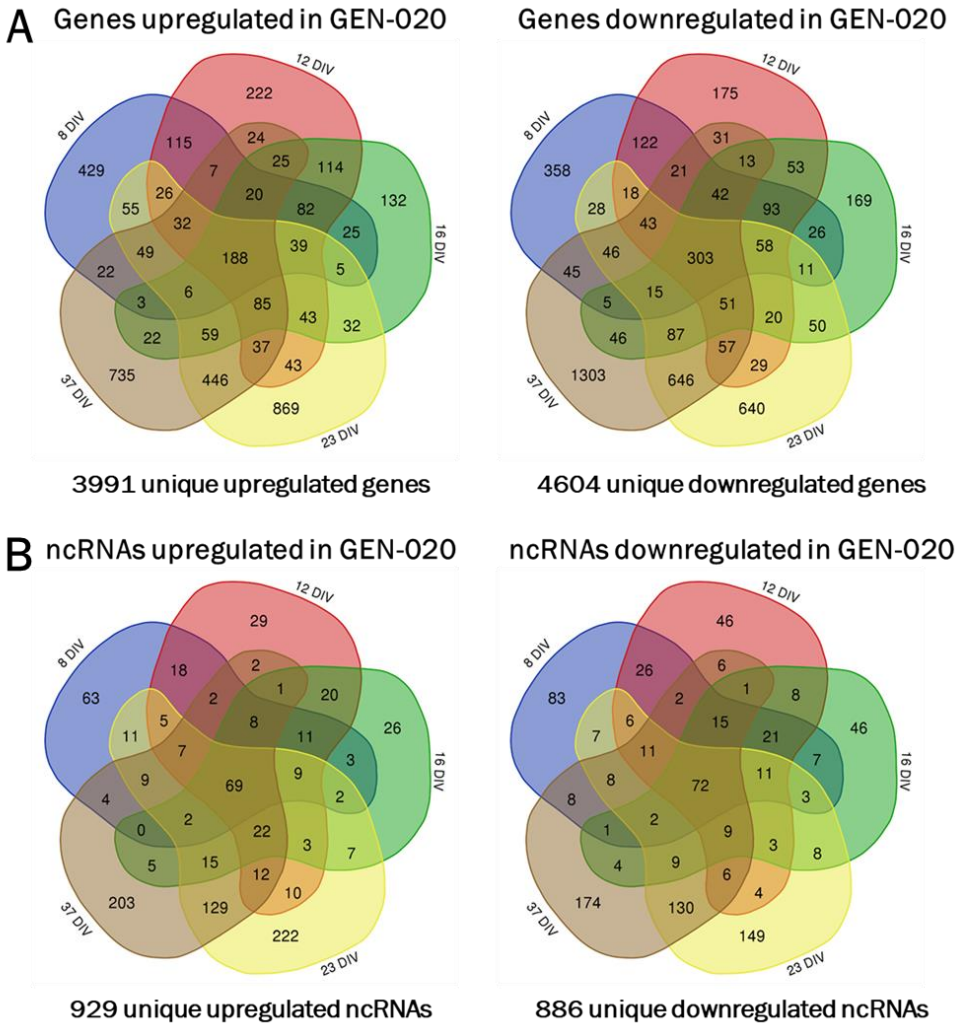


C

	0 DIV	8 DIV	12 DIV	16 DIV	23 DIV	30 DIV	37 DIV
ICC	✓	✓		✓	✓		✓
OpenArrays	✓	✓		✓	✓		✓
Single cell RNA-seq		✓	✓	✓			
Bulk RNA-seq		✓	✓	✓	✓		✓
MEAs					✓	✓	✓
Calcium imaging							✓

Figure 1. Using human pluripotent stem cell *in vitro* differentiation and a diverse range of analytical methods to study striatal development. A: Forebrain neuron *in vitro* differentiation protocol summary. **B:** Overview of analyses performed during *in vitro* differentiation. **C.** Summary of timepoints analysed with each method. OpenArray is a high throughput qPCR platform for analysing the expression of selected panels of genes involved in striatal development, and neuronal function and maturation. ICC – immunocytochemistry. MEAs – multielectrode arrays. DIV – days *in vitro*.

Results I: Non-coding RNAs are differentially expressed during neurodevelopment in HD.



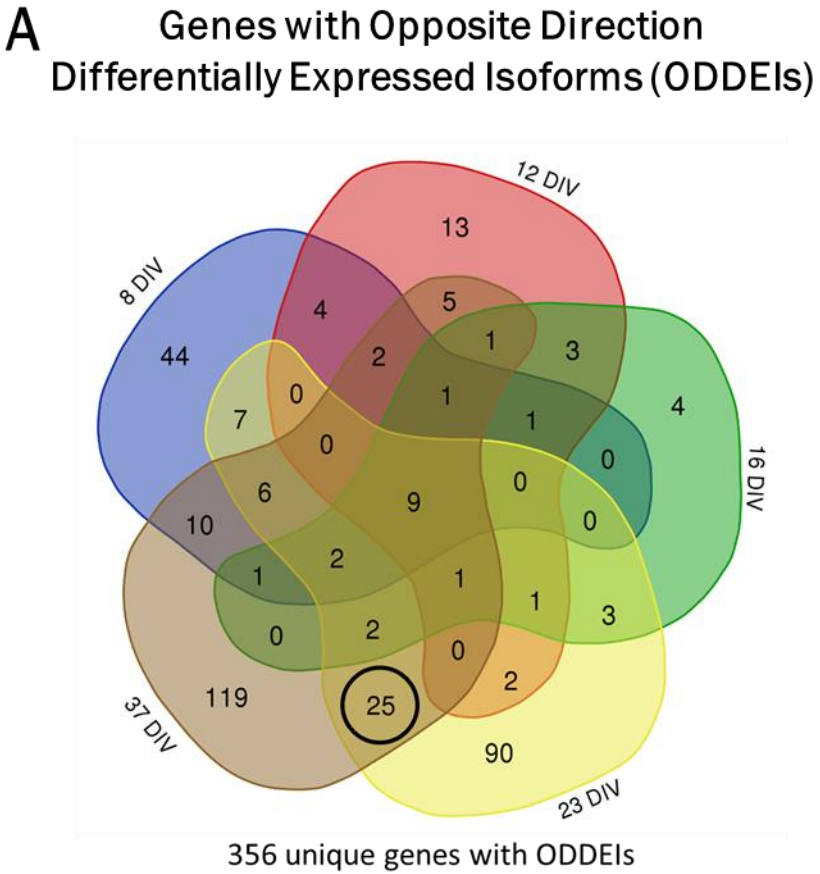
A total of 8,595 differentially expressed genes (DEGs) were identified between the two cell lines across the five timepoints. 3,991 genes were upregulated in GEN-020 (HD) compared to GEN-019 (control) while 4,604 were downregulated. 5,032 DEGs were differentially expressed at a single timepoint while the remainder were differentially expressed at two or more timepoints (Figure 2A).

We identified 1,815 differentially expressed non-coding RNAs (ncRNAs) including long ncRNAs, microRNAs and antisense RNAs (Figure 2B). ncRNAs have an important role in neurodevelopment[13] and are also dysregulated in HD[14]. The identification of these differentially expressed ncRNAs may provide insight into how neurodevelopment is affected in HD.

Figure 2. Identification of differentially expressed genes including non-coding RNAs during *in vitro* differentiation of control and HD PSC lines. Summary of all differentially expressed genes (A) and differentially expressed non-coding RNA (ncRNA) genes (B) in GEN-020 (HD) compared to GEN-019 (control) at each timepoint. Left and right – upregulated and downregulated genes in GEN-020 compared to GEN-019 respectively. DEGs have a shrunken fold change between the two cell lines > 1.5 and a false discovery rate < 0.05. n=3 for each cell line. DIV = days *in vitro*.

13. Zhang *et al* (2017). *Hum. Mol. Genet.* **26**, 3202-3211.
14. Johnson (2012). *Neurobiol. Dis.* **46**, 245-254.

Results II: Alternative splicing of neurodevelopment-related genes is altered in HD



Alternative splicing (AS) is important during neurodevelopment[15,16] and is dysregulated in HD[17-19]. We identified many differentially expressed isoforms (DEIs) between GEN-019 and GEN-020 at each timepoint (data not shown). For the majority of genes with DEIs, all the DEIs mirror the DEG, e.g. if gene X is upregulated in GEN-020 at 16 DIV, all gene X DEIs at 16 DIV are upregulated in GEN-020.

However, we identified 356 genes with at least one upregulated DEI and one downregulated DEI in GEN-020 at the same timepoint (Figure 3A). We defined these isoforms as opposite direction DEIs (ODDEIs) with some genes having ODDEIs at more than one timepoint. The presence of ODDEIs suggests that for various genes one isoform is swapped for another during GEN-020 (HD) differentiation compared to GEN-019 (control). Such changes in isoform expression are likely to affect neurodevelopment as they will alter the functional properties of the proteins that are expressed and/or their spatiotemporal expression profiles.

Figure 3. Genes with opposite direction differentially expressed isoforms during control and HD PSC line *in vitro* differentiation are associated with neurodevelopmental functions. A. Summary of genes with opposite direction differentially expressed isoforms (ODDEIs) at each timepoint during *in vitro* differentiation. Circle highlights the 25 genes with ODDEIs that are common to the neuronal timepoints (23 and 37 DIV). Differential expression is defined as in Figure 2. n=3 for each cell line. DIV = days *in vitro*.

15. Zhang *et al* (2016). *Cell.* **166**, 1147-1162. 16. Fogel *et al* (2012). *Hum. Mol. Gen.* **21**, 4171-4186.
17. Lin *et al* (2016). *Hum. Mol. Gen.* **25**, 3454-3466. 18. Fernández-Nogales *et al* (2016). *Brain Path.* **26**, 772-778. 19. Cabrera *et al* (2017). *Brain Path.* **27**, 181-189.

Results II: Alternative splicing of neurodevelopment-related genes is altered in HD

B Enriched functional terms from the 356 unique genes with ODDEIs

Term name	Adjusted p value
Cytoskeleton	1.63E-11
Axon	8.35E-09
Nervous system development	1.37E-08
Neurogenesis	3.95E-08
Neuron development	5.39E-08
Cell morphogenesis	1.83E-07
Neuron differentiation	4.49E-07
Neurodevelopmental abnormality	4.59E-02

C Enriched functional terms from the 25 genes with ODDEIs common to the neuronal timepoints

Term name	Adjusted p value
Vesicle	2.13E-04
Cytoskeletal protein binding	6.75E-04
Regulation of microtubule polymerization	9.29E-03
Cytoskeleton	5.77E-03
Postsynaptic density	1.03E-02
Neuron to neuron synapse	1.53E-02
Main axon	1.70E-02

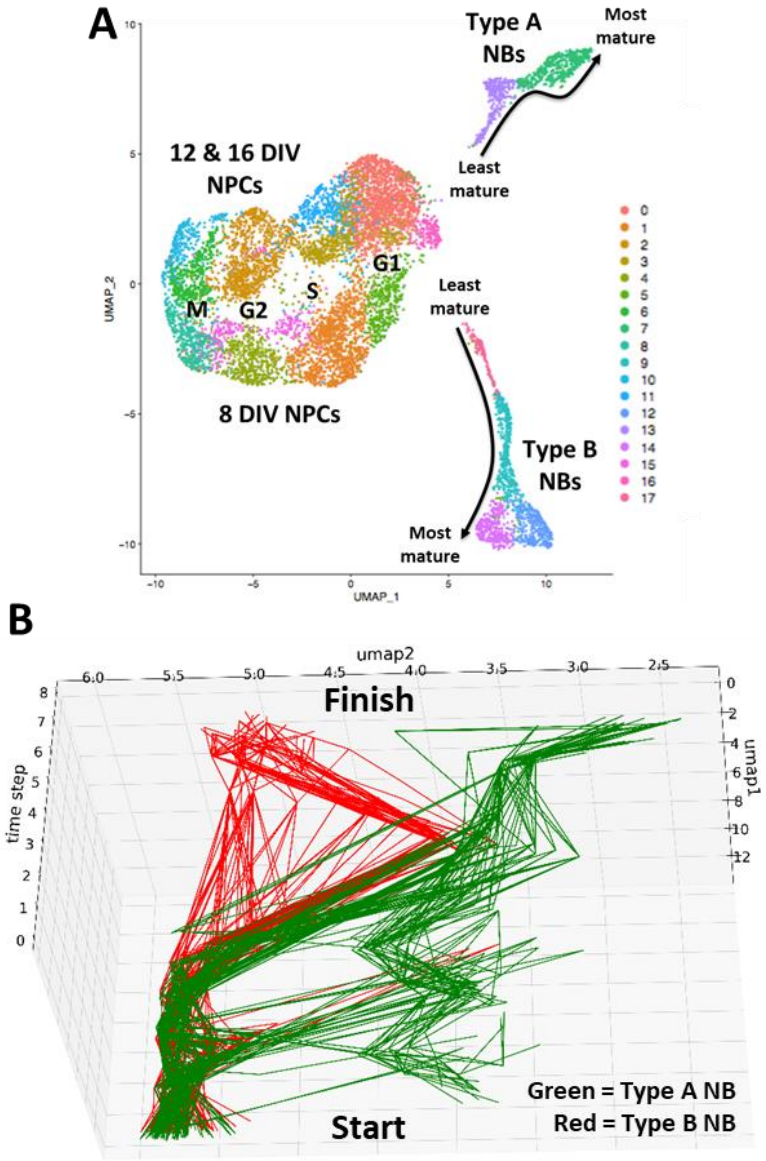
Functional enrichment analysis of the 356 genes with ODDEIs identified numerous terms associated with neurodevelopment including “neurodevelopmental abnormality” (Figure 3B). This suggests that AS is altered during neurodevelopment in HD and identifies AS as a potential target for therapeutic intervention.

As neuronal AS is affected in HD[18,19], we analysed the 25 genes with ODDEIs that are common to the neuronal timepoints (23 & 37 DIV – circled area in Figure 3A). MAP2 is among these genes and has an imbalance of low and high molecular weight isoforms that has been described previously in HD[19]. Functional enrichment analysis of the 25 genes identified various cytoskeletal and neuronal-associated terms (Figure 3C) which suggests that dysregulated AS in HD may affect neuronal morphology, function and maturation.

Figure 3 continued. Genes with opposite direction differentially expressed isoforms during control and HD PSC line *in vitro* differentiation are associated with neurodevelopmental functions. B. Functional enrichment analysis of the 356 genes with opposite direction differentially expressed isoforms (ODDEIs). A selection of enriched terms is displayed from a total of 229. **C.** Functional enrichment analysis of the 25 genes with ODDEIs common to the neuronal timepoints (highlighted with a circle in Figure 3A). A selection of enriched terms is displayed from a total of 35. Enriched terms have an adjusted p value < 0.05.

18. Fernández-Nogales *et al* (2016). *Brain Path.* **26**, 772-778. 19. Cabrera *et al* (2017). *Brain Path.* **27**, 181-189.

Results III: Neuroblast sub-types have distinct developmental trajectories

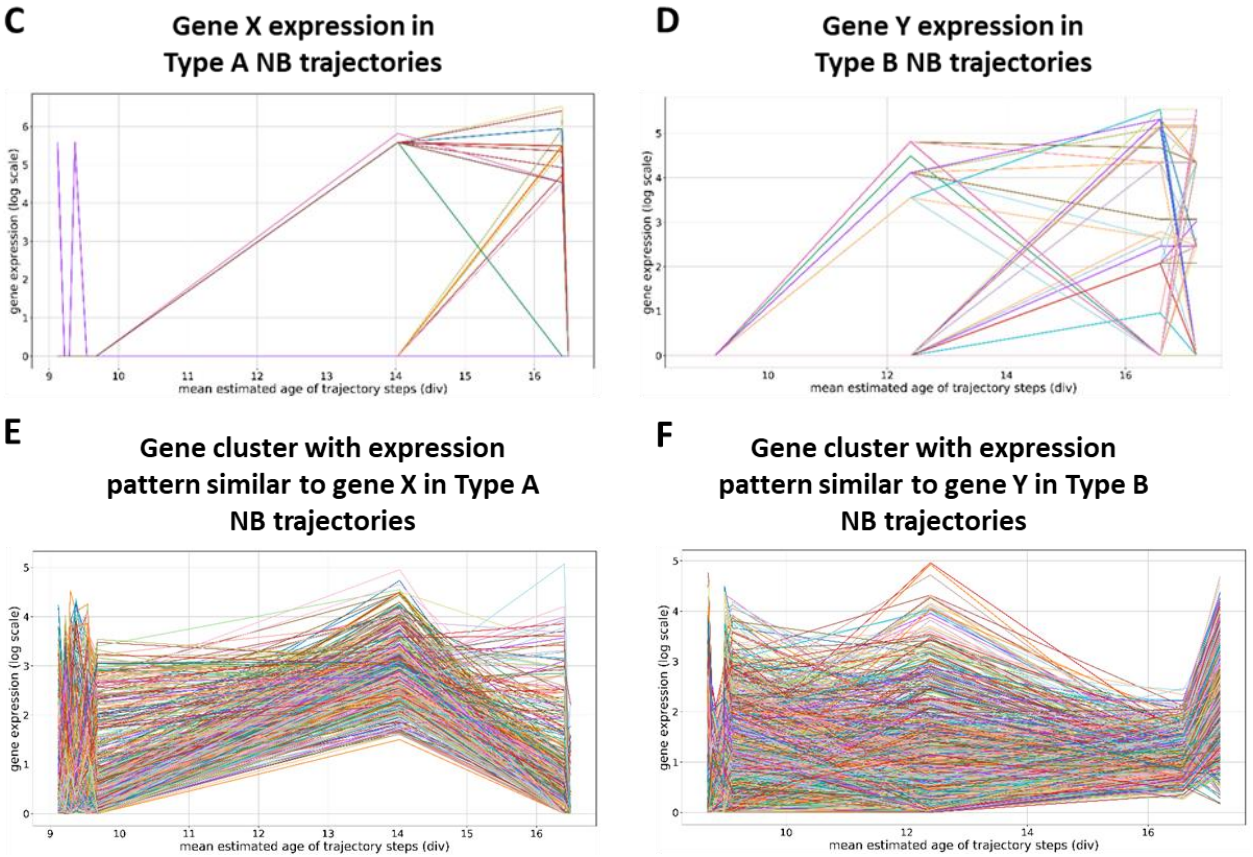


To complement the bulk RNA-seq analysis we performed single cell RNA-seq (scRNA-seq) at 8, 12 and 16 DIV. Combined analysis of the three timepoints identified neural progenitor cell (NPC) and neuroblast (NB) populations (Figure 4A). The NPCs sub-clustered according to developmental age and cell cycle phase. Two main NB types were identified which both sub-clustered according to their maturation state. However, from the clustering analysis it was unclear how the NB types developed from the NPCs.

To investigate NB development we constructed developmental trajectories using the the complete scRNA-seq data set. We identified two main trajectory routes corresponding to the Type A and B NBs (green and red in Figure 4B), and observed that for each NB type there are multiple routes to the same developmental endpoint. Despite the distinct developmental endpoints, the initial stages of a subset of Type A trajectories overlap with those of all Type B trajectories. This suggests that at least a subset of these two NB types may be derived from a common progenitor.

Figure 4. Striatal neuroblast sub-types have distinct developmental trajectories and gene expression profiles. A. UMAP analysis of the complete single cell RNA-seq data set. 18 distinct cell clusters were identified (see colour code). NPCs = Neural progenitor cells. NBs = neuroblasts. G1, S, G2 and M denote cell cycle phases. DIV = days *in vitro*. **B.** Developmental trajectories of the two NB types identified in A. Green and red lines denote the trajectories of the Type A and Type B NBs respectively. On the time step axis 0 and 8 denote the youngest and oldest cells respectively in the trajectories. Start and finish mark the beginning and end of the trajectories.

Results III: Neuroblast sub-types have distinct developmental trajectories



To investigate how gene expression varies during striatal development we plotted the expression of genes X and Y, that are involved in the specification of distinct striatal neuron sub-types, onto the Type A and Type B NB developmental trajectories. We observed distinct expression patterns in the two trajectories due to the different roles of these genes in striatal neuron sub-type specification (Figure 4C & D).

We then identified gene clusters with expression patterns similar to genes X and Y along the developmental trajectories (Figure 4E & F). These gene clusters are likely to have a role in the development of specific striatal neuron sub-types and include genes that have not previously been implicated in striatal development.

Figure 4 continued. Striatal neuroblast sub-types have distinct developmental trajectories and gene expression profiles. C & D. Expression profiles of striatal neuron sub-type marker genes X (C) & Y (D) along the Type A (C) and Type B (D) NB developmental trajectories. Gene X expression was not detected in Type B NB trajectories. Gene Y expression was not detected in Type A NB trajectories. Coloured lines denote expression along different trajectories that converge on the same endpoint. **E & F.** Expression profiles of gene clusters with expression patterns similar to gene X along the Type A NB developmental trajectory (E) and gene Y along the Type B NB developmental trajectory (F). Each line represents one gene. For each gene the expression value at each time step is the average of all of the different trajectories. In C to F the age of the different time steps was estimated in "days *in vitro*".

Results IV: Integrating data sets to generate a predictive model of striatal development

We are using bioinformatic and machine learning approaches to integrate the diverse range of data sets described here (Figure 1B & C) to generate a predictive model of striatal development. We anticipate that this model will predict:

- The gene expression profile of cell types in between available timepoints.
- How cells will develop beyond the final available timepoint.
- How a developmental trajectory will change if the expression of specific genes is altered or disabled.
- The genes that could be altered to correct HD cells to a healthy developmental trajectory.

To enhance the model we will also integrate human striatal data sets and data from the striatum of embryonic and post-natal wild type mice and HD mouse models. This will include spatial transcriptomic data to obtain valuable insight into the relevance of spatial positioning during striatal development.

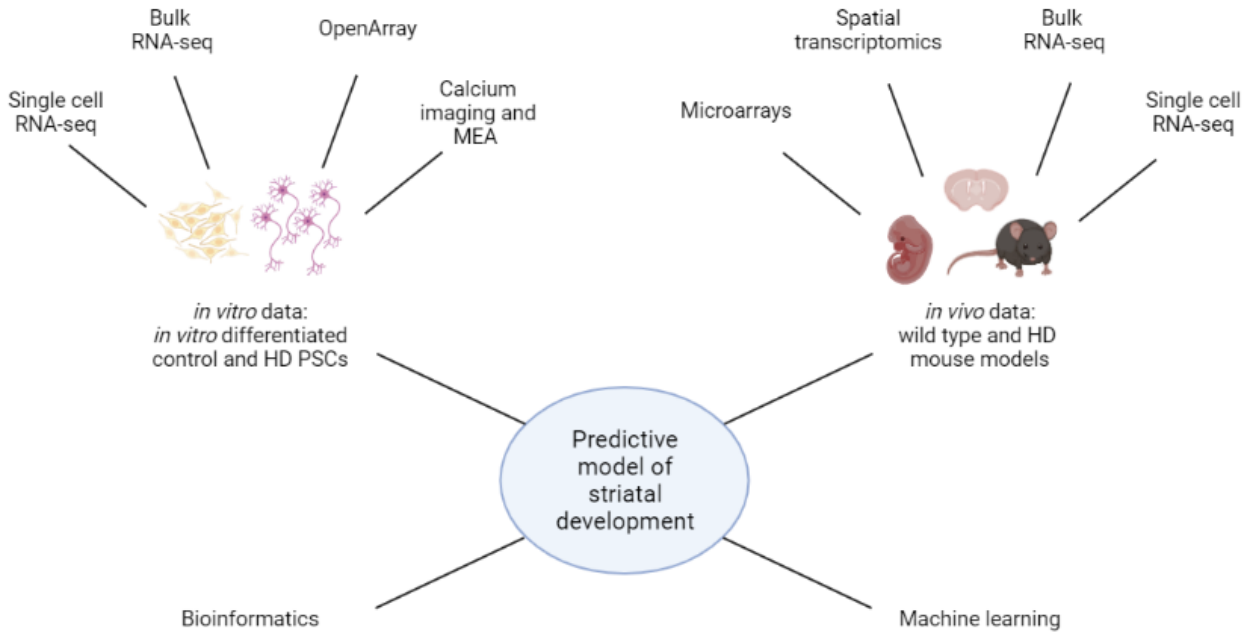


Figure 5. Towards building a predictive model of striatal development. Outline for integrating *in vitro* and *in vivo* data sets to model striatal development using bioinformatic and machine learning approaches. MEA – multielectrode array. Figure generated using Biorender.

Conclusions

1. Non-coding RNA expression is altered during HD PSC differentiation towards a MSN fate.
2. Alternative splicing is perturbed in differentiated HD PSCs and affects genes involved in neurodevelopmental processes.
3. Striatal neuroblast sub-types have distinct developmental trajectories.
4. Mapping the expression of known striatal marker genes onto the developmental trajectories identifies novel genes involved in striatal development.
5. The integration of diverse striatal development data sets will generate a predictive model of striatal development.

This study was supported by grants from the Ministerio de Ciencia, Innovación y Universidades (RTI2018-099001-B-I00); Instituto de Salud Carlos III, Ministerio de Ciencia, Innovación y Universidades and European Regional Development Fund (ERDF) (Red de Terapia Celular, RD16/0011/0012); and Generalitat de Catalunya (2017SGR-1408), Spain; and the CHDI Foundation (A16399 to J. M. C.), USA.