proceedings

MDPI

Proceedings

Heterogeneity across striatal neurons derived from patients with and without Huntington's disease generated using different *in vitro* differentiation protocols

Ziqi Li

[1] Huafu International; 13822194560@163.com
* Correspondence: 13822194560@163.com; Tel.: +86 138 2219 4560
† Presented at The 2nd International Electronic Conference on Brain Sciences, Online, 15-30 July 2021.

Abstract: Mutations in the huntingtin gene on chromosome 4 leading to the repetition of the CAG codon more than 36 times will result in autosomal dominant Huntington's disease. Mutant HTT protein not only disrupts the gene regulation of mitochondrial function-related proteins, but also affects microtubules, mitochondrial dynamic fusion and division, and calcium transport. In this research, we aimed to investigate the cellular pathways affected by mutant HTT in *in vitro* models of Huntington's disease generated with different *in vitro* differentiation protocols. To address this aim, we re-analyzed publicly available single-cell RNA-Seq datasets derived from induced pluripotent stem cells generated from patients with and without Huntington's disease. The count matrices provided by the authors were re-analyzed using the Seurat package in RStudio. The use of single-cell RNA-Seq datasets and *in vitro* models of Huntington's disease allows us to compare transcriptional signatures across neurons derived from patients with Huntington's disease generated with different differentiation protocols. The gene expression differences between neurons derived from Huntington's disease and control patients may help to shed light on the cellular pathways affected in Huntington's disease.

Keywords: Huntington's disease; iPSC; scRNA-Seq

Citation: Li, Z. Heterogeneity across striatal neurons derived from patients with and without Huntington's disease generated using different *in vitro* differentiation protocols. **2021**, *68*, x. https://doi.org/10.3390/xxxxx

Received: date
Accepted: date
Published: date

1. Introduction

Huntington's disease is an autosomal dominant neurological disorder caused by the mutation in the huntingtin (*HTT*) gene. CAG repeat expansion in the *HTT* gene of more than 36 CAG units leads to Huntington's disease [1]. Although progress has been made towards a deeper understanding of the mechanistic cause of Huntington's disease, our knowledge about the pathways and therapeutic markers is not yet complete [13].

Induced pluripotent stem cells (iPSCs) became a popular model for understanding neurodegenerative diseases in the recent years. iPSC-derived models of Huntington's disease are powerful models of the disease, that have been shown to recapitulate the disease pathology *in vitro* [6,7]. The faithfulness to which the *in vitro* models recapitulate the *in vivo* Huntington's disease pathology is yet to be determined [3].

2D and 3D in vitro models derived from iPSCs from Huntington's disease patients were previously used to understand the mechanism of Huntington's disease. Single-cell RNA-Seq using 2D model of iPSCs derived from Huntington's disease patients and unaffected controls revealed that WNT signaling is affected in Huntington's disease, and can be corrected using modulators of WNT signaling [9]. A similar scRNA-Seq dataset using iPSCs derived from juvenile forms of Huntington's disease and corrected controls suggested that there is a substantial variability in gene regulation in Huntington's disease, and this could affect mRNA regulation and polyQ aggregation [10]. Specifically, proteins that are known to physically interact with HTT are more variable in cells derived from

Huntington's disease compared to controls, which may be associated with pathology. A 3D neuroloid model of Huntington's disease was developed to overcome the variability in *in vitro* models and was found to recapitulate the basic hallmarks of human neurulation [5].

This project focused on harvesting previously generated scRNA-Seq data from iPSCs derived from patients with Huntington's disease to address the variability between different *in vitro* models of Huntington's disease. The main aim was to identify pathways commonly dysregulated across different models of Huntington's disease using publicly available datasets and to determine whether 2D and 3D *in vitro* models of Huntington's disease share common transcriptional signatures when compared to unaffected controls.

2. Methods

Publicly available scRNA-Seq data generated from iPSCs derived from Huntington's disease and matched controls were used in this project (GSE118682 [5]; GSE1444477, [9]; GSE138525 [10]).

The scRNA-Seq datasets were re-analyzed using Seurat package in RStudio [2, 4, 8, 11]. Wherever available, normalized instead of raw counts were used to construct Seurat objects, and the same parameters as reported by the authors were used for processing the data. The normalized count matrices were analyzed separately for each dataset in RStudio using the , via UMAP clustering and using the cell type markers provided by Haremaki et al., (2019, [5]). The following gene expression markers were used to identify striatal neurons of interest: PAX6, OTX2, EMX2, LHX5, IRX3, SOX10, FOXD3, ETS1, NGFR, SNAI2, SIX1, EYA2, NEUROG1, POU4F1, HES6, KRT19, ANXA1, GATA3, and CLDN6, STMN2, FAM57B, SYT4, SCG3, and GPRIN3.

Once the striatal neurons in each dataset were identified, the cells were extracted from each matrix and integrated using the IntegrateData function in Seurat [2]. The differentially expressed genes across the datasets were identified using FindMarkers function and visualized using DoHeatmap in Seurat [4].

3. Results and Discussion

The main goal of this project was to re-analyze publicly available scRNA-Seq datasets from different *in vitro* models of Huntington's disease to identify common transcriptional signatures of *in vitro* neurons with Huntington's disease. For this, we used the following data available on GEO Datasets: 1. 3D neuroloid model generated from patients with and without Huntington's disease (GSE118682, [5]), 2. 2D model of iPSCs derived from Huntington's disease patients and unaffected controls (GSE1444477, [9]), 3. 2D model of iPSCs derived from patients with juvenile Huntington's disease and corrected controls (GSE138525, [10]).

3.1. Re-analysis of GSE118682 from a 3D neuroloid model of Huntington's disease

ScRNA-Seq generated from neuroloids derived from patients with and without Huntington's disease was first re-analyzed, to established whether this pipeline can recapitulate the results seen by original authors (GSE118682, [5]).

The neuroloid datasets derived from a patient with Huntington's disease (hd) and without (ctrl) were integrated and main cell types were identified using parameters and cell type markers previously reported by the authors of this datasets [5]. Although this analysis did not filter cells based on cell cycle, main clusters of cell types were identified and the analysis recapitulated the results previously identified by the authors of this dataset [5]. The original results by the authors identified six main clusters of cell types, in contrast to our eight, reflected by two clusters for neuroepithelial and neural crest cells each instead of one (Figure 1).

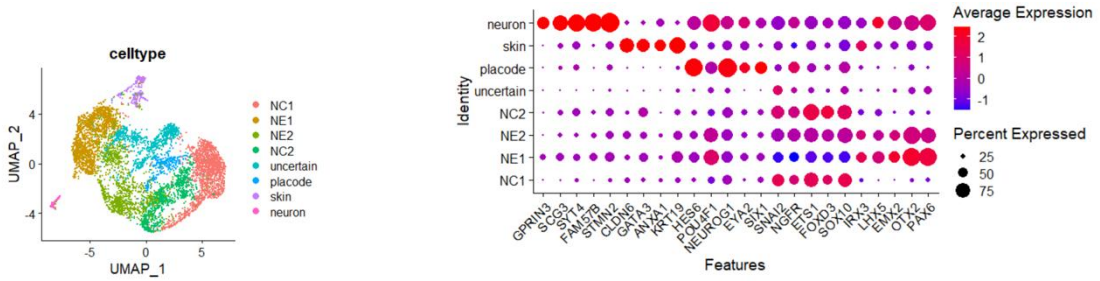


Figure 1. Re-analysis of the GSE118682 dataset [5] and clustering with UMAP revealed eight clusters, with identifiable cell identities. NC = neural crest, NE = neural epithelium.

We repeated a similar analysis using the same cell type markers on the other two scRNA-Seq datasets, GSE1444477 and GSE138525. In the dataset GSE144477, neuronal cluster was clearly separate from all other cell types using the markers of striatal neurons from Haremaki et al. (2019) [5]. The dataset GSE138525 contained 299 cells of Huntington's disease and matched control cells and contained two clusters of cell types, with unclear separation between striatal neuron subtypes.

3.2. Integrating neuronal populations from different scRNA-Seq datasets derived from iPSCs from patients with and without Huntington's disease

In the next part of the analysis, only striatal neurons from each dataset were taken forward. The neuroloid data matrix (GSE118682) was subseted to contain only the neuronal cluster. This led to the identification of 85 cells, 41 control (ctrl) and 44 Huntington's disease neurons (hd). The GSE144477 data matrix was also subseted to contain only striatal neurons, leading to the identification of 323 cells, 308 derived from the patient with Huntington's disease (53n) and 15 cells derived from the patient without Huntington's disease (18n). The GSE138525 matrix with 299 was taken forward in full, containing 68 cells derived from a patient with juvenile form of Huntington's disease (180CAG), 57 cells from a matched control (96ex), 89 cells derived from a patient with Huntington's disease (HDiPSC) and 85 cells from its matched control (C116). These cells were integrated to form a single Seurat object.

The cells derived from Huntington's disease patients and controls were renamed to 'hd' and 'ctrl', respectively. Although control and hd striatal neurons separated clearly in the UMAP plot (Figure 2), there were no differentially expressed genes (DE) between ctrl and hd neurons. The top three markers identified with FindAllMarkers function in Seurat did not reach more than 0.325 average log2FC, although the p-value was significant.

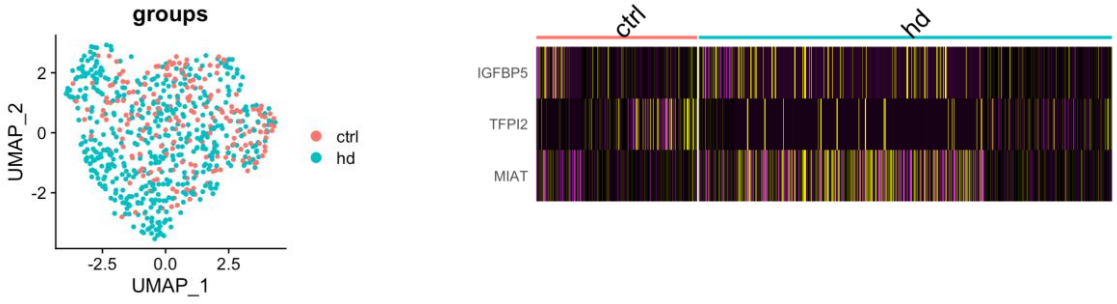


Figure 2. Integration of striatal neurons from GSE118682, GSE138525, and GSE144477 datasets reveals separation of ctrl and hd neurons, but no clear transcriptional differences between the two groups.

3.3. Integrating neuronal populations from scRNA-Seq datasets derived from 2D in vitro models containing iPSCs from patients with and without Huntington's disease

To identify whether the 3D model could introduce bias in the results, the Seurat object was subseted to include only 2D in vitro models, excluding the GSE118682 neuroloid dataset. This created a count matrix with 622 cells in total, 157 ctrl and 465 hd cells. Ctrl and hd groups separated clearly and seven DE genes were identified between ctrl and hd cells (Figure 3). However, it is clear that the 180CAG cells from juvenile form of Huntington's disease are driving this enrichment.

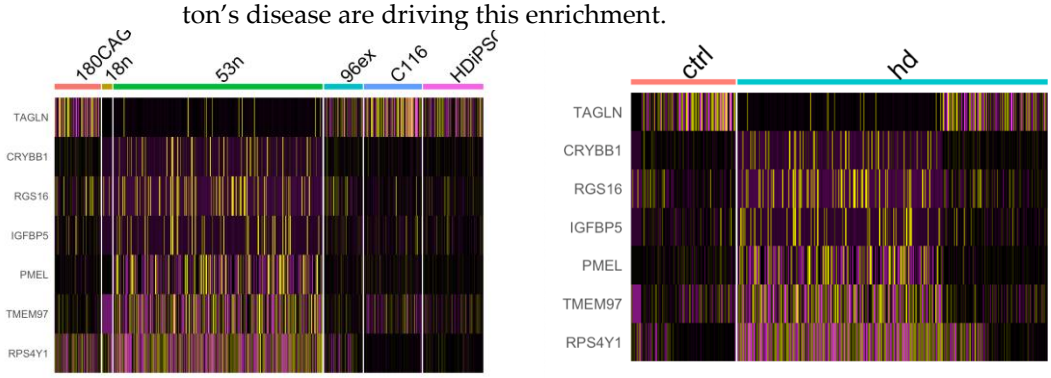
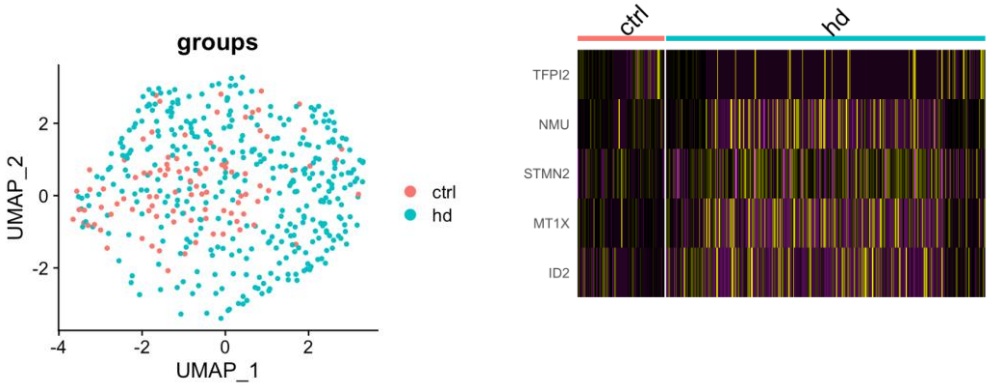


Figure 3. Integration of striatal neurons from GSE138525 and GSE144477 datasets only suggests transcriptional differences between ctrl and hd groups, but driven by a single sample.

3.4. Improved integration of neuronal populations from scRNA-Seq datasets derived from 2D and 3D in vitro models containing iPSCs from patients with and without Huntington's disease

We repeated the filtering of datasets and integration of striatal neurons, by excluding the juvenile form of Huntington's disease and its respective control (180CAG and 96ex) from the dataset. This improved our ability to identify striatal neurons in the GSE138525 dataset, 42 in the C116 sample and 45 neurons in the HDiPSC sample. These were integrated with striatal neurons from datasets GSE118682 and GSE144477 as described above. Through additional filtering and identification of markers using FindAllMarkers function, five genes were identified as DE between ctrl and hd (Figure 4). As previously, however, two samples were driving this transcriptional difference, this time from the GSE144477 dataset. The TFIP2 gene remains interesting (average log2FC 0.252, p-value 0.0000000137), as this was identified in the previous comparison in section 3.2. TFIP gene was previously implicated in extracellular matrix adhesion by neurons [12].



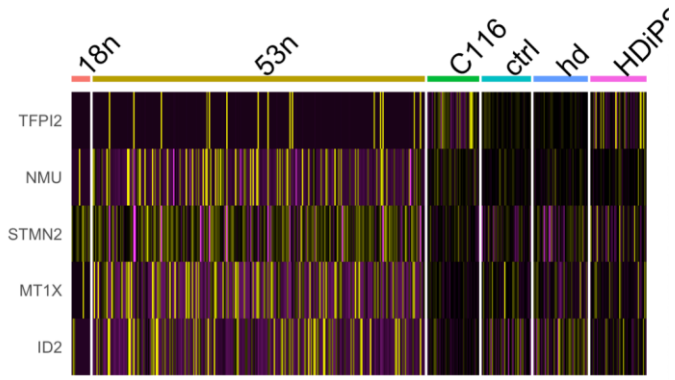


Figure 4. Improved integration of striatal neurons from GSE118682, GSE138525, and GSE144477 datasets reveals separation of ctrl and hd neurons and some transcriptional differences, driven by single *in vitro* dataset GSE144477.

3.5. Discussion

This analysis attempted to find common transcriptional signature across iPSC-derived striatal neurons from patients with Huntington's disease, using publicly available scRNA-Seq datasets using 2D and 3D *in vitro* models.

We re-analyzed three scRNA-Seq datasets, using normalized data deposited in GEO Datasets: GSE118682, GSE138525 and GSE144477. However, we could not find any shared transcriptional signatures across the striatal neurons common to neurons derived from Huntington's disease patients vs controls. Side comparisons and heatmaps plotted per individual samples, although identified DE genes, this difference in gene expression was largely driven by individual samples, and was not shared across all samples, either hd or ctrl.

Our re-analysis was likely confounded by the use of cell type markers: we chose to use the markers of striatal neurons from GSE118682 dataset published by Haremaki et al. (2019) [5]. However, these are markers of striatal neurons from 3D cultures, and further care should be taken to confidently identify all striatal neurons in 2D cultures. In addition, we did not filter cells based on cell cycle phase, which could have contributed with additional noise to our re-analysis.

4. Conclusion

In summary, we could not identify common transcriptional signatures across striatal neurons from different models of Huntington's disease compared to neurons from controls. *In vitro* models remain heterogeneous and it remains to be determined which models best recapitulate Huntington's disease *in vitro*.

Author Contributions: Z.L. carried out the research and wrote the article, S.H. provided technical support, support for conceptualization, supervision and planning, writing and review of the article.

Institutional Review Board Statement: Not applicable"

Informed Consent Statement: Not applicable"

Data Availability Statement: Data re-analyzed in this study is publicly available via GEO Database: GSE118682, GSE138525 and GSE144477

Acknowledgments: Thanks belongs to Silvia Hnatova, who supervised this project.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bates, G.P., Dorsey, R., Gusella, J.F., Hayden, M.R., Kay, C., Leavitt, B.R., Nance, M., Ross, C.A., Scahill, R.I., Wetzel, R., Wild, E.J., Tabrizi, S.J., 2015. Huntington disease. Nature Reviews Disease Primers 1, 1–21. https://doi.org/10.1038/nrdp.2015.5
2. Butler, A., Hoffman, P., Smibert, P., Papalexi, E., Satija, R., 2018. Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nature Biotechnology 36, 411–420. https://doi.org/10.1038/nbt.4096
3. Csobonyeiova, M., Polak, S., Danisovic, L., 2020. Recent Overview of the Use of iPSCs Huntington's Disease Modeling and Therapy. Int J Mol Sci 21. https://doi.org/10.3390/ijms21062239
4. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zagar, M., Hoffman, P., Stoeckius, M., Papalexi, E., Mimitou, E.P., Jain, J., Srivastava, A., Stuart, T., Fleming, L.B., Yeung, B., Rogers, A.J., McElrath, J.M., Blish, C.A., Gottardo, R., Smibert, P., Satija, R., 2020. Integrated analysis of multimodal single-cell data. bioRxiv 2020.10.12.335331. https://doi.org/10.1101/2020.10.12.335331
5. Haremaki, T., Metzger, J.J., Rito, T., Ozair, M.Z., Etoc, F., Brivanlou, A.H., 2019. Self-organizing neuruloids model developmental aspects of Huntington's disease in the ectodermal compartment. Nature Biotechnology 37, 1198–1208. https://doi.org/10.1038/s41587-019-0237-5
6. Mehta, S.R., Tom, C.M., Wang, Y., Bresee, C., Rushton, D., Mathkar, P.P., Tang, J., Mattis, V.B., 2018. Human Huntington's Disease iPSC-Derived Cortical Neurons Display Altered Transcriptomics, Morphology, and Maturation. Cell Reports 25, 1081-1096.e6. https://doi.org/10.1016/j.celrep.2018.09.076
7. Nekrasov, E.D., Vigont, V.A., Klyushnikov, S.A., Lebedeva, O.S., Vassina, E.M., Bogomazova, A.N., Chestkov, I.V., Semashko, T.A., Kiseleva, E., Suldina, L.A., Bobrovsky, P.A., Zimina, O.A., Ryazantseva, M.A., Skopin, A.Yu., Illarioshkin, S.N., Kaznacheyeva, E.V., Lagarkova, M.A., Kiselev, S.L., 2016. Manifestation of Huntington's disease pathology in human induced pluripotent stem cell-derived neurons. Molecular Neurodegeneration 11, 27. https://doi.org/10.1186/s13024-016-0092-5
8. Satija, R., Farrell, J.A., Gennert, D., Schier, A.F., Regev, A., 2015. Spatial reconstruction of single-cell gene expression data. Nature Biotechnology 33, 495–502. https://doi.org/10.1038/nbt.3192
9. Smith-Geater, C., Hernandez, S.J., Lim, R.G., Adam, M., Wu, J., Stocksdale, J.T., Wassie, B.T., Gold, M.P., Wang, K.Q., Miramontes, R., Kopan, L., Orellana, I., Joy, S., Kemp, P.J., Allen, N.D., Fraenkel, E., Thompson, L.M., 2020. Aberrant Development Corrected in Adult-Onset Huntington's Disease iPSC-Derived Neuronal Cultures via WNT Signaling Modulation. Stem Cell Reports 14, 406–419. https://doi.org/10.1016/j.stemcr.2020.01.015
10. Sorek, M., Oweis, W., Nissim-Rafinia, M., Maman, M., Simon, S., Hession, C.C., Adiconis, X., Simmons, S.K., Sanjana, N.E., Shi, X., Lu, C., Pan, J.Q., Xu, X., Pouladi, M.A., Ellerby, L.M., Zhang, F., Levin, J.Z., Meshorer, E., 2021. Pluripotent stem cell-derived models of neurological diseases reveal early transcriptional heterogeneity. Genome Biol 22. https://doi.org/10.1186/s13059-021-02301-6
11. Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., Hao, Y., Stoeckius, M., Smibert, P., Satija, R., 2019. Comprehensive Integration of Single-Cell Data. Cell 177, 1888-1902.e21. https://doi.org/10.1016/j.cell.2019.05.031
12. Swindell, W.R., Bojanowski, K., Kindy, M.S., Chau, R.M.W., Ko, D., 2018. GM604 regulates developmental neurogenesis pathways and the expression of genes associated with amyotrophic lateral sclerosis. Translational Neurodegeneration 7, 30. https://doi.org/10.1186/s40035-018-0135-7
13. Tabrizi, S.J., Flower, M.D., Ross, C.A., Wild, E.J., 2020. Huntington disease: new insights into molecular pathogenesis and therapeutic opportunities. Nature Reviews Neurology 16, 529–546. https://doi.org/10.1038/s41582-020-0389-4