R E S C I E N C E C

[Re] Spread of alpha-synuclein pathology through the brain connectome is modulated by selective vulnerability and predicted by network analysis

Thomas Paul[1,2] and Mathieu Bourdenx[1,2, ID]

[1]Univ. de Bordeaux, Institut des Maladies Neurodégénératives, UMR 5293, Bordeaux, France – [2]CNRS, Institut des Maladies Neurodégénératives, UMR 5293, Bordeaux, France

1 Introduction

Parkinson's disease (PD) is a common motor neurodegenerative disorder, characterized in part by the progressive loss of pigmented dopaminergic neurons from the *Substantia Nigra pars compacta*. The main neuropathological hallmark of PD is the progressive accumulation of α-synuclein in fibrillar aggregates named Lewy Bodies (in the soma) or Lewy Neurites (in neurites) [1]. Interestingly, Lewy pathology follows a predictable pattern of progression in the brain suggesting that α-synuclein can propagate from neurons to neurons [2]. Such observation, common to many neurodegenerative diseases [3], prompted initiatives to model and predict pathology progression in the brain of patients with neurodegenerative disorders.

In 2012, Raj and colleagues developed a Network Diffusion Model (NDM) to predict atrophy progression in dementia [4]. For this purpose, they compared Magnetic Resonance Imaging recordings from healthy young subjects, healthy old subjects, and old patients suffering from either Frontotemporal Dementia or Alzheimer's Disease. Their study suggested that patterns of atrophy within the dementia spectrum could be explained by transmission of the disease along neuronal pathways. Using their model, they demonstrated a path for predicting future atrophy in individuals starting from baseline. The NDM model was later applied to other diseases using various kind of data. In 2017, Mezias and Raj showed that the spread of amyloid-β pathology, involded in Alzheimer's disease, in mice was driven by spatial proximity but not network connectivity [5]. At the opposite, progression of tau pathology, also involved in Alzheimer's disease and other tauopathies, was better recapitulated by connectivity [6]. In 2019, Henderson and colleagues reported that progression of α-synuclein pathology, similar to tau pathology, is largely predicted by anatomical connectivity between brain structures [7].

We here replicated the core of the NDM applied to α-synuclein spread in mice as implemented by Henderson and colleagues [7]. Both R scripts, connectivity matrix and experimental datasets were made fully accessible by the authors on GitHub allowing an easy process. We were able to reproduce qualitatively and quantitatively the main results of the paper. The aim of the present replication is not to fully replicate the entire paper but to focus on the core model and main immediate results.

2 Background

2.1 Graph Theory and brain connectivity

Henderson et al. combined quantitative pathology mapping in the mouse brain with network modeling to understand the spatiotemporal pattern of spread of α-synuclein pathology [7]. Graph theory is a branch of mathematics applicable to Neuroscience. The brain is a complex system that can be modeled as a network or graph. When applied properly, Graph theory can offer important insights into different aspects of brain networks such as architecture, evolution, development, or clinical disorder [8].

Networks are defined as a collection of elements (or nodes) and their pairwise links (or edges) that can be summarized in the form of a connectivity matrix (named adjacency matrix). Depending on the type of graph, edges can have binary values (present: 1 - absent: 0) or actual weights reflecting connection strength. A graph is called undirected when an edge e reciprocally connects nodes V_a and V_b. Conversely, a graph is named directed when the edge e is projecting from V_a to V_b but not from V_b to V_a.

Thanks to the initiative from the Allen Institute for Brain Science's Mouse Connectivity Atlas (MCA), the mouse brain connectome is now freely accessible for the neuroscience community [9]. In a matrix, connectivity is represented as "outgoing" along rows and "incoming" along columns. From the adjacency matrix, the in-degree and out-degree distributions can be computed as respectively the sum of all entries in the corresponding row and the sum of all entries in the corresponding column and are represented in diagonal matrices.

The Laplacian graph is a matrix used to explore the properties of a network. It is computed using the adjacency matrix and either the in-degree or out-degree graph depending on the applications.

2.2 Model description

The model used by Henderson et al. requires the mouse brain connectome (discussed in the previous section) and whole brain α-synuclein pathology quantification.

The model used in the study is the classical α-synuclein pre-formed fibrils (PFFs - 5μg) unilateral injection in the dorsal striatum of non-transgenic mice (NTG) [10, 7]. Following inoculation, NTG mice were terminated at 3, 6, or 9 months post-injection (MPI). Brain pathology was assessed using traditional immunohistochemical methods to detect pathological α-synuclein phosphorylated on Serine129 (pS129). Quantification (as percent of region occupied by immunostaining) was performed on both brain hemispheres (ipsilaterally and contralaterally to the injection point) in 58 different regions. Thus, the pathology dataset provided by the authors on their repository is an Excel sheet with values for 5 mice per group and timepoint for all 116 brain regions.

Using the dataset from [9] for synaptic connection, the authors generated a directed and weighed connectivity graph $G = V, E$ whose nodes V are N cortical and subcortical grey matter regions and whose edges $e_{ij} \in E$ represent an axonal projection from V_i to V_j. They then defined the weighed adjacency matrix of G as $A = [A_{ij}]$ generating a final parcellation of 116 regions.

The magnitude of **observed** α-synuclein pathology of all N nodes at a time t is the vector x_t. The **predicted** regional α-synuclein pathology $\widehat{x_t}$ is a function of the adjacency matrix A and a seed region $seed$ ($\in E$) and is computed as shown in **equation 1**:

$$\widehat{x}_t = e^{-cLt} x_o \tag{1}$$

Where:

- c is a constant designed to achieve an optimal match with the data as the empirical diffusion constant for the pathology is *a priori* unknown.

- t stands for the timepoint of the prediction.

- L is the out-degree Laplacian matrix computed as shown in (2):

$$\begin{cases} -A_{ij} & \text{for } i \neq j \\ \sum_{j=1}^{N} A_{ij} & \text{for } i = j \end{cases} \tag{2}$$

- x_0 is the seed vector computed as shown in (3):

$$x_0 \begin{cases} 0 & \text{for } i \neq seed \\ 1 & \text{for } i = seed \end{cases} \tag{3}$$

3 Material and Methods

3.1 Replication of the Model

Henderson et al. coded this model using RStudio 3.3.3. Their code and data are accessible on GitHub. Two major folders are located in their repository. The first one contains the scripts that are specific to their analysis. The second folder contains the numerous data sets used in the study and notably the pathology data set, the connectivity matrices, and the *Snca* gene expression energy value data set.
Using Python 3.7 and the Operating System Windows 10, we successfully replicated the scripts summarized in the **Table 1**. [We also ran the scripts with Mac OS **[version]** to look for inter-compatibility between Operating Systems.

Table 1. Scripts Successfully replicated

From Henderson et al.	Equivalent in Python	Function
process_pathconn.R	process_files.py	Processes the input data sets
getLout.R	Laplacian.py	Compute the out-degree Laplacian matrix
fitfxns.R → make.Xo	fitfunctions.py	Sets the seed of the model
fitfxns.R → predict.Lout	fitfunctions.py	Predicts the pathology
fitfxns.R → c.fit	fitfunctions.py	Obtains a c value that fits the model

We successfully replicated the core of Henderson et al. code. subsequently, we were capable of computing the pathology prediction $\widehat{x}$(t) for the three different timepoints in NTG mice. We attempt to reproduce the figures related to the main code from the original article.

3.2 Pipeline of our code

Our code consists of a main script named pipeline.py made of a DataManager to easily change the different parameters such as the timepoints, the seed, or the quantified pathology matrix. *pipeline.py* successively calls the scripts mentioned below and saves

the data as tables and figures.

process_files.py processes the data inputs (Quantified pathology matrix, ipsilateral connectivity matrix, and contralateral connectivity matrix). Its main outputs are the adjacency matrix, a pathology matrix, and the list of regions of interest. Both matrices were shaped in a panda DataFrame.

Laplacian.py computes, from the adjacency matrix, the out-degree Laplacian matrix. After careful observations, we successfully replicated the Laplacian matrix that captures the retrograde connectivity.

fitfunctions.py is a set of functions with different roles. *make_Xo* has been reproduced as Henderson's code. We shaped it as a column DataFrame that contains 0 for non-seed regions or 1 for the seed region. *predict_Lout* predicts the regional α-synuclein pathology when given the Laplacian, the seed, a timepoint of prediction, and the tuning constant c. *c_fit* returns the value of the best fit c and the best correlation coefficient using *predict_Lout* and Pearson's correlation test.

The pipeline also calls different scripts that we used for our own study (stability of the model, robustness, the iterative model). More information on how to run the code and the required structure of the inputs can be encountered in the ReadMe section available in our corresponding GitHub. **Add an hyperlink when done**

4 Results

4.1 Network model of pathological α-synuclein spread

The network diffusion model validated by Henderson et al. is based on anatomical connectivity. Using this model, they estimated pathology as a function of time given the introduction of misfolded α-synuclein into the ipsilateral Caudoputamen (iCPu).
Figure 2.A illustrates the correlation between our predicted data and the quantified data. The first observation between our replicated predictions and the original ones is a visual correspondance between our Scatterplots and those displayed by Henderson et al. We can then appreciate Pearson's r scores equal to those of Henderson and colleagues such as 0.56, 0.69 and 0.65 with respect to MPI 1, MPI 3, MPI 6. Our model also computes p-values that are equal to those from the replicated article before and after Bonferroni correction. **Table 2** summarizes these results. These data suggests that linear dynamics imposed on a connectivity network can explain pathology spread over time.

Then, we imposed to our model some controls to assess its robustness. As performed in the original code, we randomly seeded the diffusion model and plotted the fit r for each timepoint. **(Figure 2.B)** This revealed that CPu seed produces either the best fit (3 MPI and 6 MPI) or the second-best fit (1 MPI) as observed by Henderson et al. as well. Likewise as published in 2017 by Pandya et al., we randomly shuffled the connectivity matrix **(Figure 2.C)** and the pathology matrix **(Figure 2.D)**. For each shuffle, we plotted the corresponding fit. **[10 Pandya, 2017]** Our results confirm that the spatiotemporal spread is indeed driven by connectivity.

4.2 Differential vulnerability of regions is correlated with α-synuclein expression

As intrinsic vulnerability had been hypothesized to be a critical factor in the development of pathology, Henderson et al. thought to investigate that aspect using the connectivity-based model. They defined the difference between predicted pathology and observed pathology as a measure of relative vulnerability. Regions with greater

Table 2. Data correspondence between the reproduced results and the original article
The results have been rounded up to the third decimal.

		Henderson et al.	Our Model
Best constant c computed		1.625	1.625
MPI 1			
	p-value before Bonferroni correction	$2.695e^{-9}$	$2.695e^{-9}$
	p-value after Bonferroni correction	$8.085e^{-9}$	$8.085e^{-9}$
	correlation R	0.559	0.559
MPI 3			
	p-value before Bonferroni correction	$< 2.2e^{-11}$	$1.150e^{-17}$
	p-value after Bonferroni correction	$3.451e^{-17}$	$3.451e^{-17}$
	correlation R	0.696	0.696
MPI6			
	p-value before Bonferroni correction	$8.793e^{-15}$	$8.793e^{-15}$
	p-value after Bonferroni correction	$2.638e^{-14}$	$2.638e^{-14}$
	correlation R	0.648	0.648

values or lower values compared to the predicted data are considered respectively as relatively vulnerable or resilient regions. **Figure 3.A** replicates the relative vulnerability computed by Henderson et al.

To determine whether the prediction could be improved, Henderson et al. seek out for factors that would explain regional vulnerability. They postulated that *Snca* gene expression should influence the development of the pathology. Using *Snca* expression energy values obtained from the Allen Brain Atlas, it was possible to investigate the relationship between *Snca* gene expression and regional vulnerability. **Figure 3.B** illustrate the *Snca* Expression energy values in different sections of the brain. In support with their hypothesis, *Snca* expression was highly correlated with vulnerability values. Both Figure A and B are displayed using the Python packageBrainrender. [**11. Claudi, 2020**]

Incorporation of *Snca* expression into the network diffusion model was performed by weighing the outgoing connections of every node by its respective *Snca* expression value. **Figure 4** is a replication of the model using the data from Henderson et al GitHub repository. Taken all together, these data suggest that *Snca* gene expression is an important factor impacting the vulnerability of regions and subsequently pathology spread. **Table 3** displays a data correspondence between the reproduced results and the original article.

4.3 In silico seeding of alternative regions

To examine the generalizability of the model, Henderson et al. proceeded to *in silico* seeding of alternative regions such as the Piriform Cortex (Pir) and the Substantia Nigra (SN). **Figure 4** display the pathology spread at MPI 1, MPI 3 and MPI 6 after seeding the model with either the Pir or the SN. We successfully obtained the same data sets of
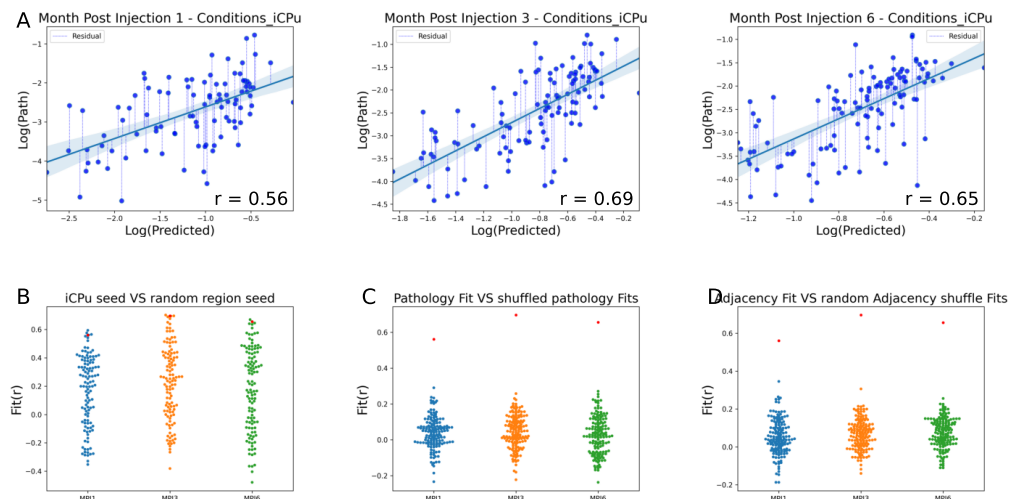
Figure 1. Network diffusion model based on anatomical connectivity explains pathological α-synuclein spread

A, Scatterplots and Pearson's correlation coefficient (r) of log(predicted) pathology based on anatomical connectivity versus actual log(pathology) values. Predictions versus pathology are shown after injection of PFF into the ipsilateral CPu for each region at a different timepoint. **B**, Each different brain region was seeded and pathology was modeled from each site. The fit of each of these sites was plotted for each timepoint. The red dot represents the fit obtained when the model is seeded into the iCPu. **C**, The pathology matrix was randomly shuffled and the fit was subsequently computed for each timepoint. The red dot represents the fit obtained when the model is seeded into the iCPu and the pathology matrix is not randomly shuffled. **D**, The connectivity matrix was randomly shuffled and the fit was subsequently computed for each timepoint. The red dot represents the fit obtained when the model is seeded into the iCPu and the adjacency matrix is not randomly shuffled.

predictions after seeding in either the Pir or the SN. *In silico* injection in the Pir remarkably corresponds to previous semi-quantitative pathology grading study as mentioned by Henderson et al. **[12. Rey, 2016]** While on the other hand, *in silico* injection in the SN resulted in slow propagation through the Caudoputamen to cortical region as observed in the staging of human cases. Taken all together, these findings support the ability of the Diffusion Network model to generalize data prediction.

5 Conclusion

In this article, we successfully replicated the model created by Henderson et al. using Python.

The Network Diffusion Model developed by Henderson et al. is a powerful tool in understanding the spread of α-synuclein pathology. Consequently, we could unveil intrinsic processes that may explain the patterns of disease spread in patients. This NDM utilizes two factors: anatomical connectivity and endogenous α-synuclein expression. Prediction of α-synuclein spread over time has revealed islands of vulnerable neurons susceptible to developing pathogenic protein inclusions. Similarly, *Snca* gene expression showed a similar pattern to inferred vulnerability. Finally, seeding in different regions such as the Pir and the SN supports the NDM generalizability. This model is attractive as it is simple, easily handled, and replicable.

Nonetheless, some limitations still need to be considered. Future empirical data are still required. Firstly, to limit sparseness in the regions sampled. Secondly, to deeply
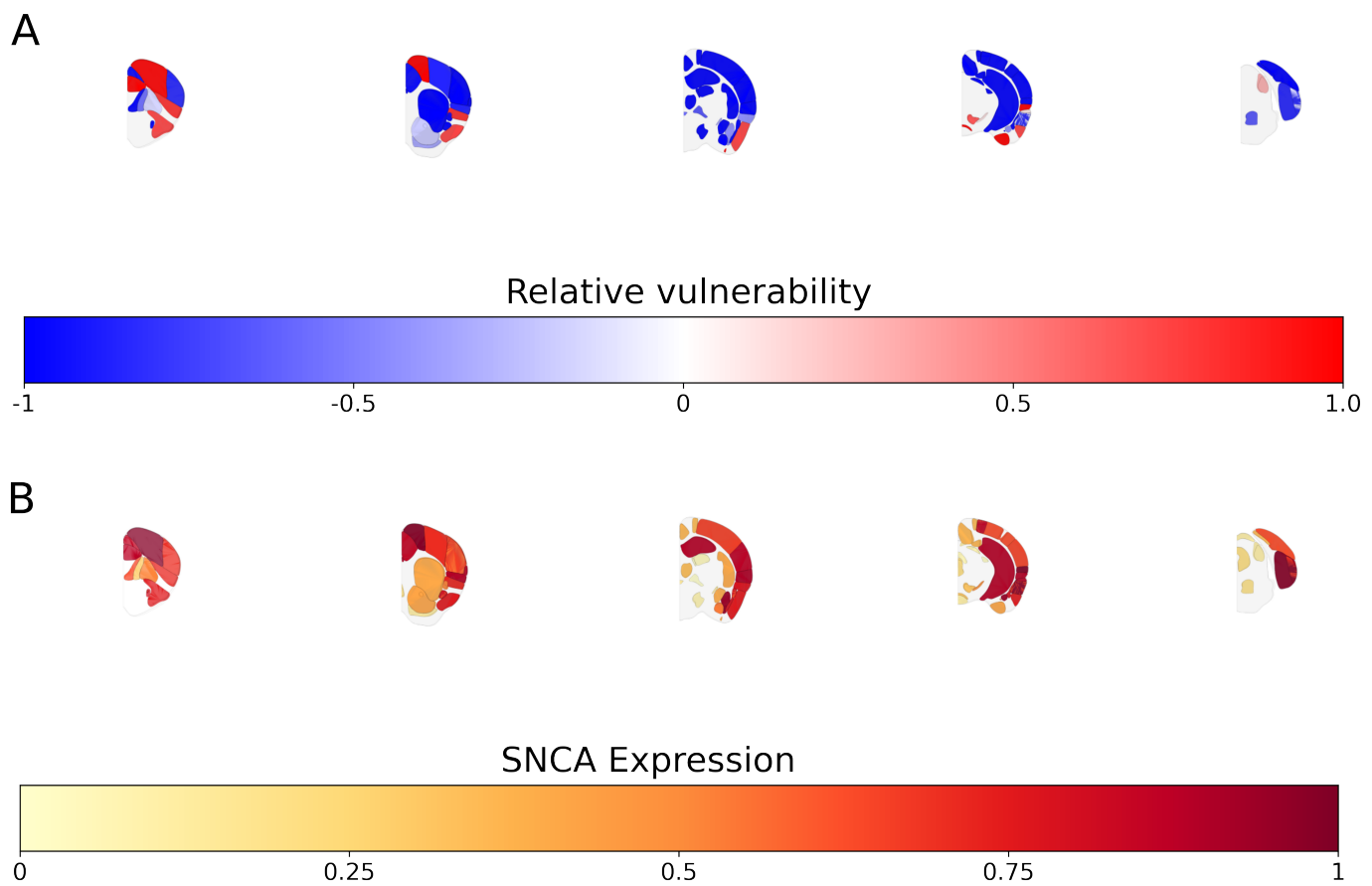
A

Relative vulnerability

-1   -0.5   0   0.5   1.0

B

SNCA Expression

0   0.25   0.5   0.75   1

Figure 2. Differential vulnerability of regions is correlated with α-synuclein expression
A, Heatmap of the residuals between the log(predicted) and log(pathology) are plotted on an anatomical mouse brain as a measure of the relative vulnerability of regions. **B**, Heat map of the SNCA mean expression energy values from the Allen Brain Atlas in situ hybridization for each of the designated region. Both Figure A and B are displayed using the Python package *Brainrender*. [11. Claudi, 2020]
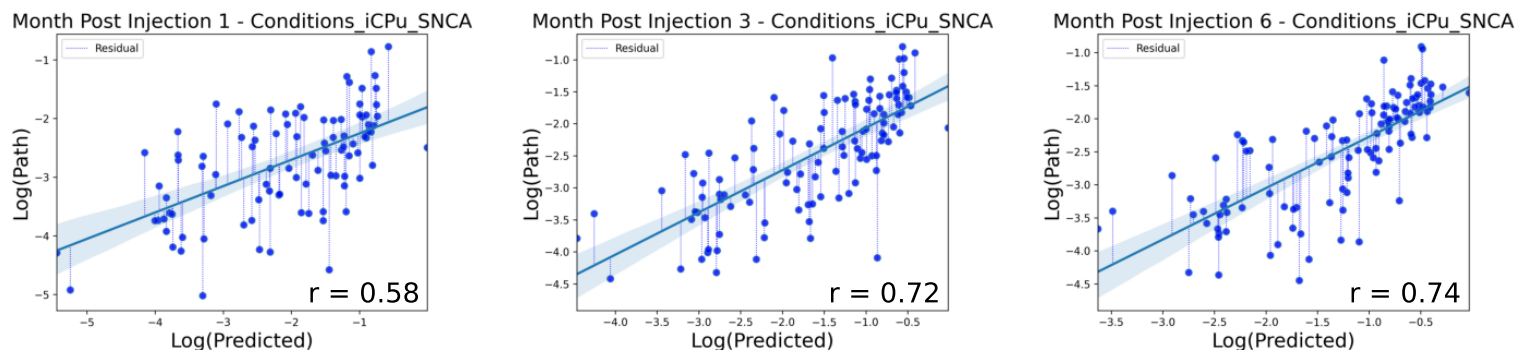
Figure 3. Incorporating *Snca* expression in the Diffusion Network Model improved the prediction
Scatterplots and Pearson's correlation coefficient (r) of log(predicted) pathology based on anatomical connectivity versus actual log(pathology) values when incorporating *Snca* expression. Predictions versus pathology are shown after injection of PFF into the ipsilateral CPu for each region at a different timepoint.
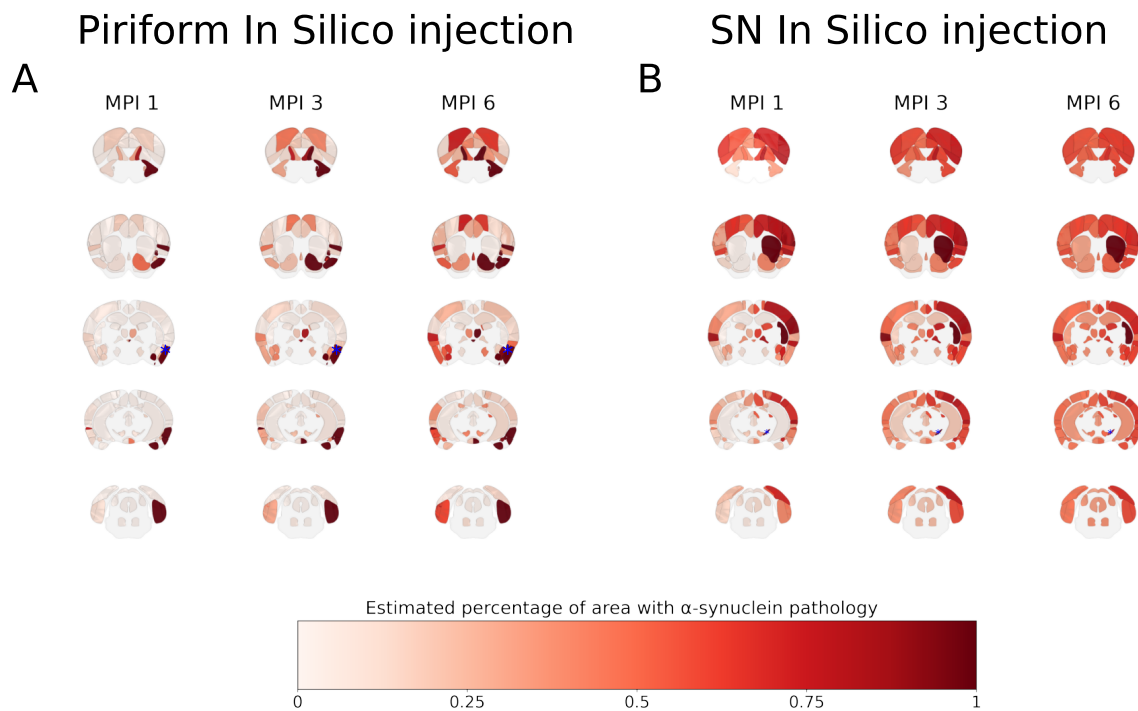
Figure 4. *In silico* seeding of alternative regions in the mouse brain
Heat maps of regions affected by α-synuclein pathology. Estimated percentage of area with α-synuclein pathology after *in silico* seeding in the Piriform Cortex (**A**) or the Substantia Nigra (**B**) The injection site is indicated on each slice by a blue asterisk.

Table 3. Data correspondence between the reproduced results and the original article after incorporation of the *Snca* expression
The results have been rounded up to the third decimal.

	Henderson et al.	Our Model
Best constant c computed	0.414	0.414
MPI 1		
p-value before Bonferroni correction	$4.67e^{-10}$	$4.67e^{-10}$
p-value after Bonferroni correction	$1.401e^{-9}$	$1.401e^{-9}$
correlation R	0.580	0.580
MPI 3		
p-value before Bonferroni correction	$< 2.2e^{-16}$	$8.071e^{-20}$
p-value after Bonferroni correction	$2.421e^{-19}$	$2.421e^{-19}$
correlation R	0.727	0.727
MPI6		
p-value before Bonferroni correction	$< 2.2e^{-16}$	$9.882e^{-21}$
p-value after Bonferroni correction	$2.965e^{-20}$	$2.965e^{-20}$
correlation R	0.739	0.739

assess the generalizability of this model. Another barrier comes from that in their study, Henderson et al. adopted a mesoscale of analysis. Their observation of a distinct laminar distribution of the pathology in cortical regions suggests that further investigation should provide data about layer and neuron types affected by the disease.

Interestingly, other models predicting pathology spread exist, such as the Spatial Diffusion Model [7. Mezias Rey,2019] or the Epidemic Spreading Model [13. Vogel et al., 2020] In their article from 2020, Henderson et al. took it even further by applying a multiple linear regression on the NDM to propose a model that examines both retrograde and anterograde contributions to the disease. [3. Henderson, 2020]

Modeling disease spread in the brain is challenging. As one could try to simulate a specific aspect, but no one would be able to simulate the brain in its entirety. Raj et al. mentioned in their review that the brain can change according to two processes that they named the "Dynamics On Network" and the "Dynamics Of Network". The first one is related to processes that occur atop a static structure as for instance prion spreading. The second defines the brain network changes over time. Parkinson's Disease exhibits dynamics in both "on" and "of" network. [14. Raj, 2018] Then, a major challenge in modeling such dynamics remains in discovering both these "on" and "of" aspects. As one could try to understand one dynamic only if the other one is not explained then we are still a step away from the reality of the pathology.

References

1. B. Dehay et al. "Targeting α-synuclein for treatment of Parkinson's disease: mechanistic and therapeutic considerations." In: **Lancet Neurol** 14.8 (Aug. 2015), pp. 855–866.
2. H. Braak, K. Del Tredici, U. Rüb, R. A. de Vos, E. N. Jansen Steur, and E. Braak. "Staging of brain pathology related to sporadic Parkinson's disease." In: **Neurobiol Aging** 24.2 (2003), pp. 197–211.

3. M. Jucker and L. C. Walker. "Propagation and spread of pathogenic protein assemblies in neurodegenerative diseases." In: **Nat Neurosci** 21.10 (Oct. 2018), pp. 1341–1349.

4. A. Raj, A. Kuceyeski, and M. Weiner. "A network diffusion model of disease progression in dementia." In: **Neuron** 73.6 (Mar. 2012), pp. 1204–1215.

5. C. Mezias and A. Raj. "Analysis of Amyloid-β Pathology Spread in Mouse Models Suggests Spread Is Driven by Spatial Proximity, Not Connectivity." In: **Front Neurol** 8 (2017), p. 653.

6. C. Mezias, E. LoCastro, C. Xia, and A. Raj. "Connectivity, not region-intrinsic properties, predicts regional vulnerability to progressive tau pathology in mouse models of disease." In: **Acta Neuropathol Commun** 5.1 (Aug. 2017), p. 61.

7. M. X. Henderson, E. J. Cornblath, A. Darwich, B. Zhang, H. Brown, R. J. Gathagan, R. M. Sandler, D. S. Bassett, J. Q. Trojanowski, and V. M. Y. Lee. "Spread of α-synuclein pathology through the brain connectome is modulated by selective vulnerability and predicted by network analysis." In: **Nat Neurosci** 22.8 (Aug. 2019), pp. 1248–1257.

8. O. Sporns. "Graph theory methods: applications in brain networks." In: **Dialogues Clin Neurosci** 20.2 (June 2018), pp. 111–121.

9. S. W. Oh et al. "A mesoscale connectome of the mouse brain." In: **Nature** 508.7495 (Apr. 2014), pp. 207–214.

10. K. C. Luk, V. Kehm, J. Carroll, B. Zhang, P. O'Brien, J. Q. Trojanowski, and V. M. Lee. "Pathological α-synuclein transmission initiates Parkinson-like neurodegeneration in nontransgenic mice." In: **Science** 338.6109 (Nov. 2012), pp. 949–953.