Translating brain anatomy and neurodegenerative disease from mouse to human through latent gene expression space

Authors: Chloe Jaroszynski[1]*, Mohammed Amer[2], Antoine Beauchamp[3], Jason P. Lerch[1,4], Stamatios N. Sotiropoulos[2,6], Rogier B. Mars[1,5]*

Affiliations:

[1]Wellcome Centre for Integrative Neuroimaging, Centre for Functional MRI of the Brain (FMRIB), Nuffield Department of Clinical Neurosciences, University of Oxford; Oxford, United Kingdom

[2]Sir Peter Mansfield Imaging Centre, School of Medicine, University of Nottingham; Nottingham, UK

[3]Holland Bloorview Kids Rehabilitation Hospital; Toronto, Ontario, Canada

[4]The Hospital for Sick Children, Canada; Mouse Imaging Centre, Canada; Department of Medical Biophysics, University of Toronto; Canada

[5]Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen; Nijmegen, The Netherlands

[6]NIHR Nottingham Biomedical Research Centre, Queen's Medical Centre, Nottingham University Hospitals (NUH) NHS Trust; Nottingham, UK

*Corresponding author. Email: chloe.jaroszynski@ndcn.ox.ac.uk; rogier.mars@ndcn.ox.ac.uk

One Sentence Summary: We predict human anatomical and disease phenotypes from the mouse brain using a translational deep learning model.

Abstract: The mouse model is by far the most widely used animal model in preclinical neuroscience, but translating findings to humans suffers from the lack of a formal framework establishing the correspondence between the mouse and the human brain. In this study, we build on the concept of common brain space, and on previous work embedding gene expression profiles, to bring the two species into alignment for comparative analysis. Using a variational autoencoder (VAE) combined with a latent classifier, we create a latent space that strikes a balance between abstract features related to reconstruction and features pertaining to regional segregation. This approach demonstrates improved cross-species homology and within-species locality compared to existing comparative models. In addition, we show that brain alterations in mouse disease models can be translated to humans, predicting patterns of brain changes in Alzheimer's and Parkinson's diseases. The flexibility and scalability of this approach offer a promising framework to bridge between more animal models, comparing quantitative imaging modalities, and disease phenotypes. This in turn will help advance our understanding of species similarities and differences, enhancing both fundamental translational neuroscience and disease modelling.

Introduction

The goal of translational neuroscience is to take results obtained in the lab and use them to improve clinical practice. Often these results are obtained in non-human models, with the mouse being the most widely used. As examples where the mouse model was instrumental in leading to new treatments in humans, one can think of the schizophrenia treatment risperidone (1), gene therapy for spinal muscular atrophy (2), or the use of monoclonal antibodies in multiple sclerosis (3). Today, preclinical research is producing more data, with more diversity and precision, than ever before. Genetic engineering techniques such as CrispR/Cas9 have enabled faster, more reliable manipulation of the mouse genome for creating models of human disease (4–6). Humanized mouse models have advanced in vivo investigation of human-specific responses to disease or treatment (7,8). However, despite the increase in sophistication of the models used and the methods to manipulate model genetics and compounds, the success rate of translation from preclinical research is low, with the average rate of success from first-in-human testing to regulatory approval at around 10% (9).

The low success rate of translational neuroscience is likely to be due to several factors, including issues in experimental design and reporting bias (10). However, one prominent cause is that the mouse and human lineages diverged over 85 million years ago (11) and both have evolved since then. As a result, the human brain differs from that of the mouse in size, morphology, and internal organization. Although both brains have been mapped using a variety of methods, resulting in detailed atlases at the molecular (12–14), cellular (15), and areal levels (16,17), formal comparison between the two species is rare. At present, there is a lack of formal comparative neuroscientific models that allow the brain of the model species and the human to be compared in a like-for-like fashion. Ideally, quantitative translational neuroscience should be able to achieve the two following goals. First, the formal comparison of the human and model species brains. Second, the ability to translate brain maps, such as the effects of disease or treatment, from one species to the other.

Simple geometry-driven registration of the mouse to the human brain is unlikely to be successful, due to the large differences in brain morphology across the two species. An alternative approach is to describe both brains in terms of a common organization feature that is comparable (18). Based on the notion of the connectivity fingerprint, which suggests that brain regions are characterized by a unique set of connections to the rest of the brain, work comparing primate brains used homologous white matter tracts as a common space to compare different species' brains. This has allowed between-species translation of cortical maps and comparisons of changes during phylogeny and ontogeny (19,20). The mouse brain, however, does not feature many dominant white matter bundles. Although functional connectivity with known homologous target areas has

been used instead (21), a different feature might be more suitable for whole-brain mouse-human comparisons.

Spatial patterns of expression of ortholog genes are a promising feature, since these have been shown to differentiate different parts of the brain (22) and transcriptomic datasets from both species are available (12,13). However, the number of features – i.e. genes – and the level of shared information across these features means that, to make the problem tractable, a reduction of the data space is necessary. We used such an approach in our earlier work (23), by training a multi-layer perceptron (MLP) to describe the mouse brain in terms of the profiles of expression of ortholog genes. The MLP's hidden layers in effect created a latent gene expression space in which both the mouse and human brain could be described, in a way that was superior to using the high dimensional gene expression space. This work provided the foundation for quantitative, whole brain comparison of the mouse and human brain based on cross-species embeddings.

While the MLP provided a latent space rich in features pertaining to regional discrimination, it did not by design enforce cross-species alignment: the cross-species similarity lacked specificity across broad anatomical areas such as the cortex and the cerebellum. One way to make the latent space more flexible, and therefore to enable adjustments meeting the specific needs of the cross-species comparison, was to switch to a variational autoencoder (VAE). The VAE, by definition, established a balance between the clustering and scattering of the distribution of data in its latent space. By simultaneously training a classifier as the latent space was being built, we additionally ensured that the clustering would respect regional boundaries, rather than any other abstract feature of the data. While highly clustered latent data would likely result in mouse and human data clustering into distinct groups in the latent space, highly scattered data might lead to loss of regional clustering within species. By adjusting the weights in the VAE loss function we optimised the model specifically to bring the two species into alignment, with the added benefit of preserving within-species regional organisation.

To compare the two species, we embedded human and mouse gene expression data into the VAE latent space and computed a latent similarity matrix which served as the basis for translation. We first performed a series of anatomical cross-species translations, identifying brain regions that translated across the two species on the basis of the similarity of their latent gene expression profiles. We subsequently use the latent space to perform two quantitative phenotype translations. We focused on two examples of neurodegeneration, Alzheimer's and Parkinson's diseases, in which we predict effects of the disease in the human based on maps obtained in the preclinical mouse model.

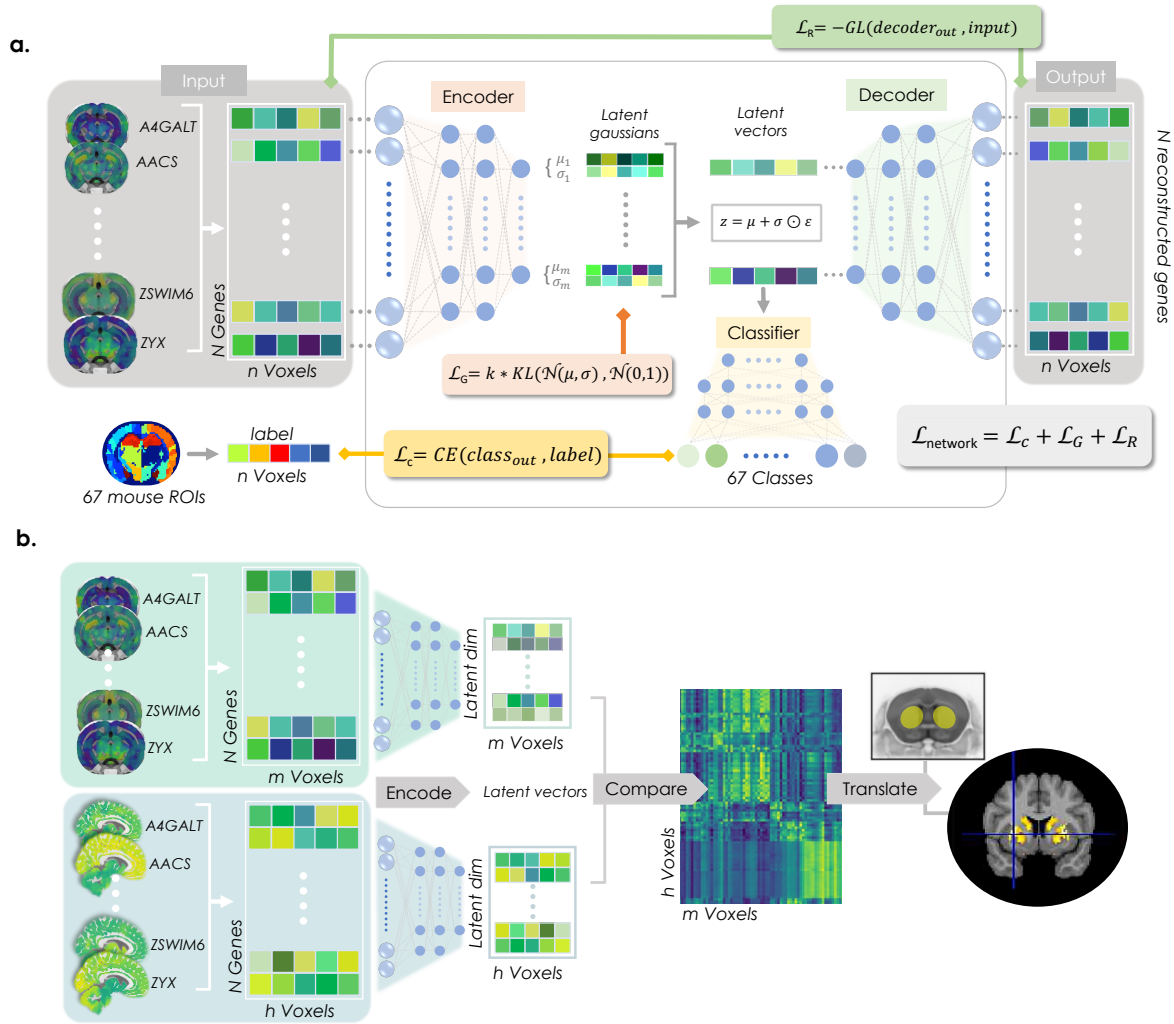

Fig. 1. Model architecture and cross-species embedding for comparison. A – Tripartite model architecture with encoder, decoder and latent classifier trained on the mouse data using N=2835 ortholog genes; GL: Gaussian likelihood; KL: Kullback-Leibler divergence; CE: cross-entropy. **B** – Application: the encoder is used to embed the mouse and human data into the latent space of the model for similarity computation and cross-species translations.

Results

1. A common space based on embedded gene expression profiles captures anatomical cross-species brain similarities.

Using spatial gene expression data (12,13) for *N*=2835 genes common across the mouse and the human, we trained an artificial neural network to learn low dimensional representations of the mouse transcriptomics data. Specifically, we used a variational autoencoder (VAE) architecture tasked with reconstructing the input data based on the latent representation of the input, and a latent

classifier tasked with recognising the regions of the Allen Mouse Brain Atlas based on the latent representation of the mouse gene expression data (Fig. 1 (A)). The joint training of the classifier and the VAE conferred additional regularisation of the latent space. After training, the encoder part of the network was used to project both mouse and human spatial transcriptomic data into the latent space (Fig. 1 (B)). This transformation enabled voxel-wise gene expression profiles – originally represented as *N*-dimensional vectors – to be compressed into *M*-dimensional latent vectors (*M*<<*N*) for both species. By establishing this common transcriptomic space, the *M*-dimensional latent representations facilitated direct comparisons and cross-species translations.

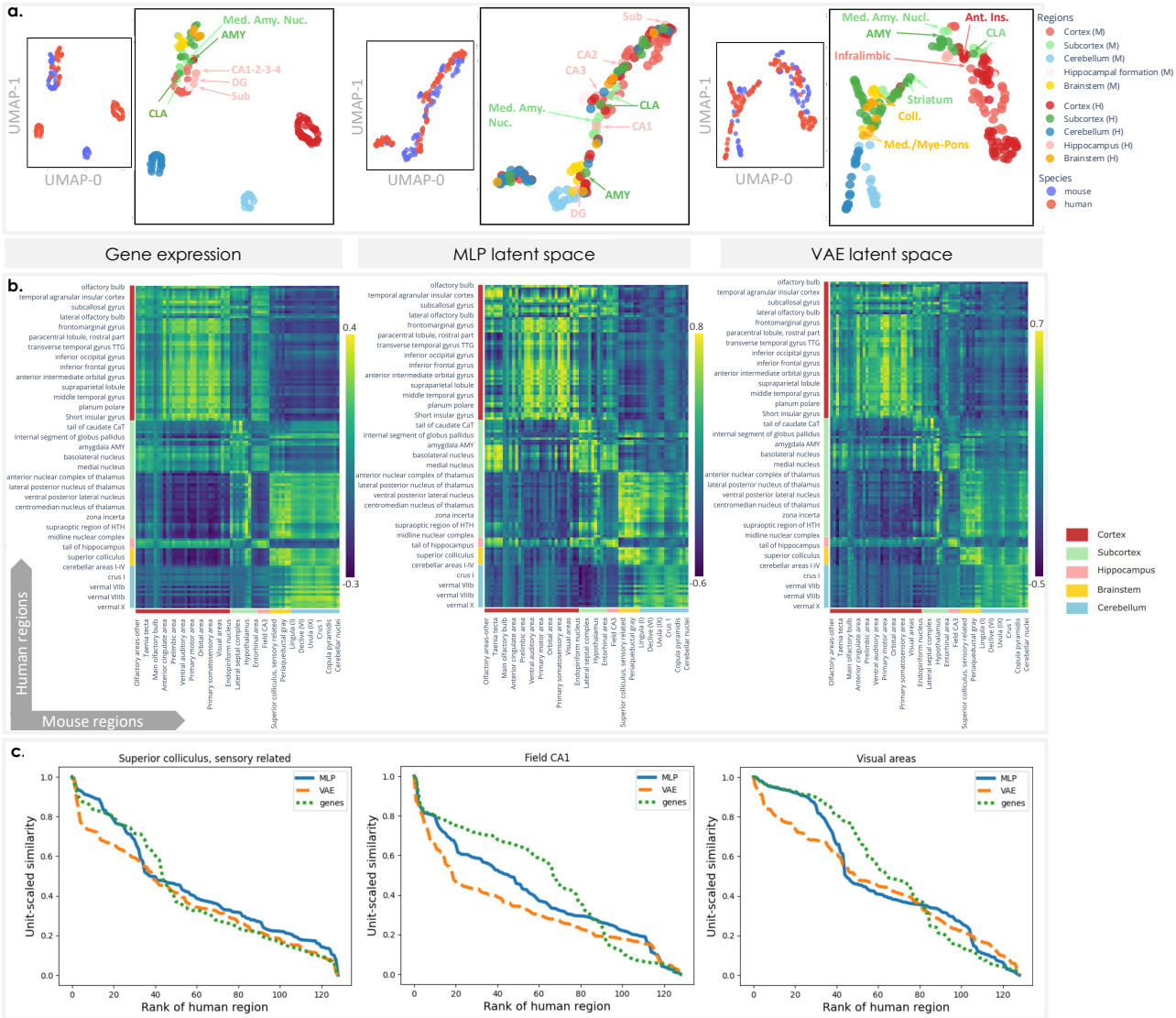

Fig. 2. Comparison of the VAE latent embedding approach with the gene expression and the MLP latent embedding approaches. A – 2-D Umap for visualisation of regional data distributions, where each dot represents a mouse or human ROI, as defined per the atlases described in the methods. The VAE achieves the best cross-species alignment and regional clustering within

species. DG=Dentate gyrus, Sub=subiculum, med. amyg. nucl=medial amygdalar nucleus, Cla/CL=claustrum, Med=medulla, mye=myencephalon, ant. ins = anterior insula. **B** – Heatmaps of region-wise cross-species similarity where axes are anatomically ordered according to major brain subdivisions (human and mouse regions on vertical and horizontal axes respectively). **C** – Unit-scaled, rank ordered similarity profiles for 3 regions and for each method. The VAE displays a steeper slope of the curve at the origin, and a lower rank at 0.75 similarity, highlighting the improvement of the VAE in terms of narrowing the similarity profile for these regions.

1.1. Alignment of cross-species distributions in 2D-Umap projections

We compared the mouse and human brain in the common transcriptomic space. As a first step we show the regions of the two species' brains together in a two-dimensional UMAP projection (24). We do this for the original high dimensional gene space, for the MLP-based approach used in our previous work (23), and for the VAE-based low-dimensional latent representation of the data.

In original gene space (Fig 2 (A**)** – left), the clustering of the data shows limited overlap between the two species. We note that the human dentate gyrus, subiculum, and hippocampal subfields clustered more closely with the mouse hippocampal formation than with other cortical regions of the same species, and that the human claustrum clusters more closely with the mouse claustrum than with other subcortical areas of the same species. The same goes for the human amygdala which clusters more closely with the mouse medial amygdalar nucleus than with other human subcortical structures. However, for all cerebellar regions, all cortical regions, most subcortical and brainstem regions, within-species clustering is predominant over cross-species homologous clustering.

For the MLP, the two-dimensional representation (Fig 2 (A) – middle) shows that in the latent space learned during training the MLP to distinguish between brain areas, the distance between species is reduced, resulting in an overlap between the mouse and the human data. However, we can also notice the loss of within-species clustering, with a mixing of cerebral, cerebellar, and subcortical/brainstem areas. This mixing is more predominant in the human than in the mouse. For instance, the human hippocampal formation is spread out instead of tightly clustered. Human field CA3 clusters closely to mouse hippocampal formation areas but is directly surrounded by human cerebellar and cortical areas. The human amygdala clusters more closely with areas of the mouse brainstem, human cortex, and human cerebellum, rather than with other areas of the human subcortex or mouse amygdalar nuclei, as was the case with the gene expression distribution. The human and mouse claustrum have remained in close proximity, although further apart than in the gene expression distribution. The loss of within-species clustering predominantly for the human may be explained by the fact that the MLP is trained on the mouse data and could therefore be less sensitive to specificities of human brain areas.

In the VAE latent two-dimensional UMAP representation (Fig 2 (A) – right), the encoded data display both cross-species overlap and within-species clustering of major anatomical subdivisions. The VAE thus appears to be less affected than the MLP to a mouse-only training and subsequent application to human data. Here, the hippocampal subfields cluster closely within and across species and are surrounded by human and mouse limbic structures. While most cortical and cerebellar areas display higher within-species clustering than cross-species clustering, brainstem and subcortical areas are well aligned across, as well as within species. This demonstrates that the VAE latent space embedding presents an improvement in cross-species alignment that, importantly, preserves anatomical segregation.

1.2. Regional cross-species similarity and improvement testing

To further investigate the improvement of regional alignment of areas across mouse and human brains using the VAE, we computed for each approach the cross-species regional similarity matrices using Pearson correlation (Fig. 2 (B)). We tested the improvement using a Kolmogorov-Smirnov test on the distribution of the similarity values across approaches and additionally investigated the differences in terms of rank ordered similarity profiles of each region (Fig 2 (C)).

The heatmaps highlight the progression of the cross-species comparison between the three approaches. On one hand, while the large high similarity clusters in the gene similarity matrix (Fig. 2 (B) – left) reflect the major anatomical subdivisions, they do not identify specific cross-species homologs. Mouse cortical regions are mostly similar to human cortical regions but have high similarity to the hippocampus and to subcortical areas; the mouse subcortex is mostly similar to the human subcortex but also to areas of the brainstem; the mouse cerebellum is mostly similar to the human cerebellum, brainstem and subcortex.

The clusters of high similarity in the MLP similarity matrix are less uniform than the gene similarity matrix, reducing the number of cross-species matches for any given region. However, matches still occur across major anatomical subdivisions, between cortex and subcortex, brainstem and subcortex, cortex and hippocampus, or cerebellum and subcortex. This is consistent with the poorer alignment observed in the MLP UMAP representations in Figure 2 (A).

In contrast to the gene and MLP matrices, the VAE similarity matrix is both more focal and anatomically accurate. Mouse cortical regions no longer have high similarity to regions outside the human cortex, mouse subcortical similarity also remains within the human subcortical areas, and the same can be seen for hippocampal fields and cerebellum. Additionally, the number of high similarity correspondences for potential cross-species homologs is further reduced.

On a general standpoint, the Kolmogorov-Smirnov test confirmed that the distributions of similarity values in the matrices were significantly different (p=5.77e-6). More specifically, the rank ordered, unit-scaled similarity profiles highlight the level of improvement for each region. A steep decay rate indicates a focal cross-species similarity, while a flat decay profile indicates that the region is highly similar to many regions in the other species. Figure 2 (C) illustrates this in 3 regions where the VAE shows steeper decay rates compared to the gene expression and MLP approaches. Looking at all 67 regions, the difference between the ranks of each profile at a 0.75 similarity level was on average in favour of the VAE by 2.7 percent or over 3 regions. Comparing the improvement of the MLP versus genes (4.63%) and the VAE versus genes (7.38%), on average the VAE showed a 60% improvement compared to the MLP.

Taken together, these observations demonstrate the efficacy of the VAE in representing regional data from both species using a common, low-dimensional latent space, with a balance between cross-species alignment and within-species clustering of brain regions.

2. Regional translations: retrieving known homologs and exploring new correspondences

2.1. Translation between known homologs

Having established the validity and specificity of the VAE latent space, we used it to compare the mouse and human brain. As a starting point, we investigated the translatability of specific regions of the mouse brain.

We first investigated the cross-species translation of the hippocampus, caudoputamen, thalamus, and hypothalamus. These areas are known to be close in anatomical space, i.e., they are located closely together in the brain but are far apart in terms of gene expression (12). Figure 3 (A) presents a translation of each of these regions, in which a binary mouse ROI, defined using the Allen Mouse Brain Atlas, was translated to human (mouse ROI multiplied by the mouse-human similarity matrix to create human brain similarity map of the mouse region, scaled to [0 1] range and thresholded as described in methods). All four regions match best to their human homologs with high selectivity, demonstrating that these regions translate well using the VAE latent space.

We next investigated a wider range of areas across all major subdivisions of the brain. Figure 3 (B) shows the translations of mouse Crus I (a cerebellar area), the pons (a brainstem area), the visual cortex (a primary cortical area), and the frontal pole (an association area of the cortex). All areas translate well in terms of having high areas of similarity in the homologous part of the human brain. However, in some cases, high similarity is also found in areas that are outside the direct homolog. For areas where this is the case, additional voxels with high similarity scores tend to be found in the same constituent parts of the brain (e.g., cortex, subcortex, cerebellum, brainstem).

This is particularly the case for the neocortical areas. For both the visual cortex and the frontopolar area, voxels with high similarity are found outside the target area but still inside the neocortex.

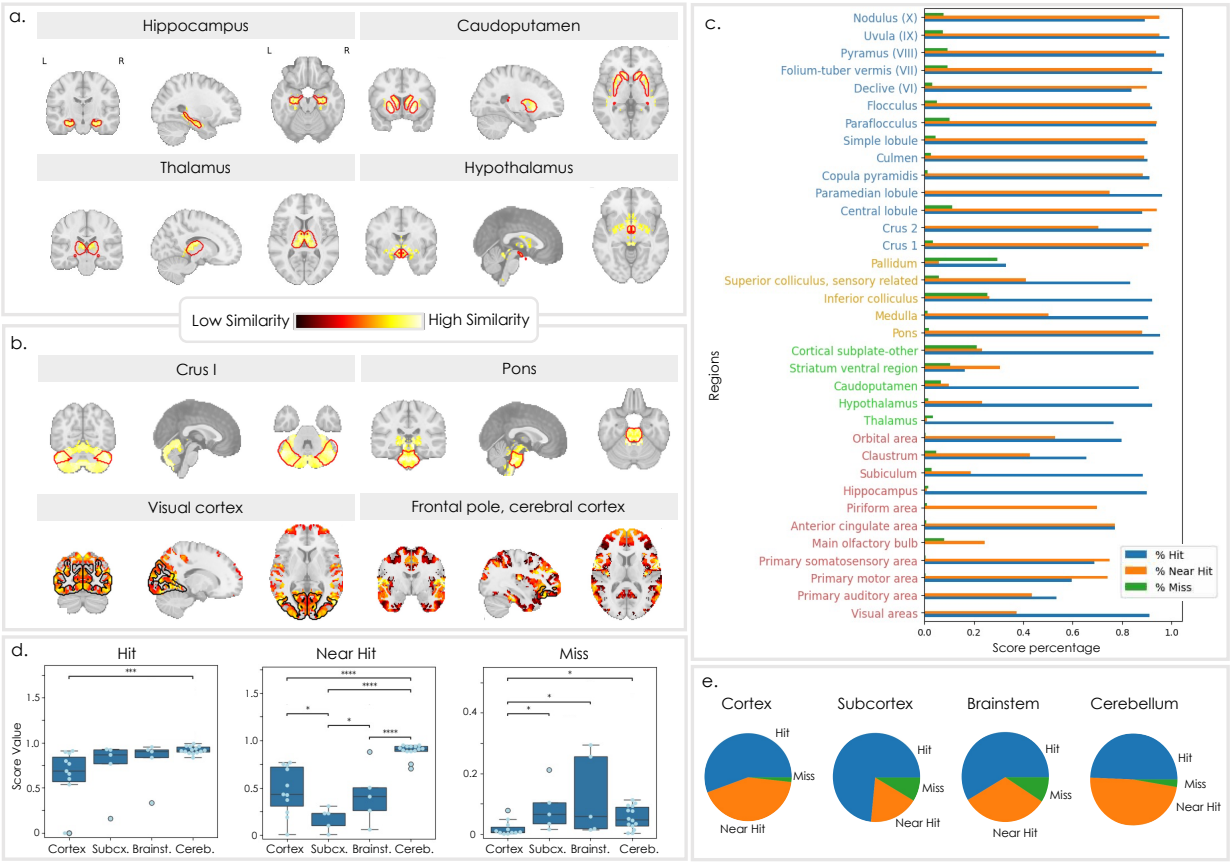

Fig. 3. Cross-species translation between homologs. A – thresholded maps of translations from hippocampus, caudoputamen, thalamus, and hypothalamus. Left and Right hemispheres are indicated on the first plot and are the same for all plots. The red line indicates the target human area. **B** – thresholded maps of translations from crus 1, pons, visual cortex, and frontal pole, with the human target indicated by a continuous red or black line. **C** – proportions of the target region containing voxels with high similarity in the translation of the mouse homolog (% hit), proportion of the larger anatomical structure containing high similarity voxels (% near hit), proportion of the rest of the brain containing high similarity voxels (% miss), for 35 homologous regions. **D** – significant differences between hit, near hit, and miss rates in the different constituent parts, according to connover's post-hoc test. **E** – proportion of score types in each constituent part.

To quantify this notion, we used predefined pairs of mouse-human homologous areas based on the literature (23), and also defined anatomical constituents for each brain region as the main anatomical area it belonged to (cortex, subcortex, brainstem, cerebellum). We then calculated for each mouse area a "hit" score, or the proportion of voxels within the homologous human area that had a high similarity (similarity is considered 'high' above threshold as defined in methods), the

proportion of voxels within the rest of the constituent part that had a high similarity ("near hit"), and the proportion of voxels in the rest of the brain that showed a high similarity ("miss"). Figure 3 (C) shows these three scores for the 35 homologous pairs. Across the whole brain, the rate of voxels falling beyond the major anatomical constituent is low. However, the pattern of scores differs across the different constituent parts of the brain. The more focal translations are found mainly in the cortex, subcortex, and brainstem, potentially highlighting a better homology between species in those areas. In 14 regions across these constituent parts, there is a higher rate of hits than near hits, and hardly any misses, which models how the similarity decreases as the distance to the target increases. In the cerebellum, the pattern is rather that there are similar proportions of hits and near hits, meaning that the translations may not be discriminating between the cerebellar subregions as well as for other constituent brain areas. An ANOVA on hit, near hit, and miss percentages for each brain area with factors SCORE TYPE (hit, near hit, miss) and CONSTITUENT PART (cerebellum, brainstem, subcortex, cortex) shows a significant interaction ($F(6,93)=10.85$, $p<0.001$). Connover's posthoc test highlights the significant pairwise differences for each score type (Fig. 3 (D)), while the pie-plots in Fig 3 (E) highlight the relative proportion of each score type, each constituent part taken independently.

The fact that the subcortex and brainstem highlight the most focal mouse-human translations is consistent with the idea that these areas are conserved in the mammalian lineage. One could have expected the same in the cerebellum, but the score pattern in the cerebellum does not necessarily mean that the translations are poor in this area. Perhaps the cerebellum is overall more different between species than the brainstem and subcortex, and would benefit from a dedicated analysis. Another reason could be that the anatomical atlas used as first intention in this study may not be the best suited for cross-species translations, and that the similarities could lie within functional boundaries (such as defined by (25)) instead.

Similarly, the variability in hit and near hit scores in the cortex should not necessarily be taken to mean that translation is worse in this part of the brain. The neocortex has expanded dramatically in the primate lineage and contains many more distinct areas in the human than in the mouse. While the pattern of organisation can be similar, the expansion that took place in the human brain results in subsequent diversification of areas (26). Cortical areas resulting from brain expansion do not necessarily have corresponding diversity in their patterns of gene expression, meaning that they could be classified as similar to mouse areas, even if they have no direct homologs in the

mouse. To further investigate this issue, we focused on the translation of cortical areas in more detail.

2.2. Neocortical translations

First, we investigated the translation of modality-specific primary areas. We showed previously that the visual areas translate accurately across species, with the highest percentage of hit voxels among primary cortices (Fig. 3 (C), while primary motor, primary somatosensory, and primary auditory areas displayed a much higher rate of off-target voxels (near hits). There is, however, some structure to the location of these translated voxels. To illustrate this, Fig. 4 (A-C) shows the annotated translations of these areas on the human brain, together with the outline of the Allen Human Brain atlas regions (27).

In the translation from the mouse primary motor area, high similarity in the human precentral gyrus is found mainly in two areas. The area on the medial surface of the cerebrum (Fig. 4 (A), top panel) is responsible for foot and toe movement, while the area adjacent to the sylvian fissure (Fig. 4 (A), bottom panel) is responsible for face, eye, lip and tongue movement. The mouse primary motor area also displays high similarity with the posterior cingulate, precuneus, and superior parietal lobule, with the lateral occipital gyrus and the angular gyrus, and with the frontal pole, middle and superior frontal gyrus, and the posterior temporal gyrus.

The translation of the mouse primary somatosensory area broadly overlaps with the translation of the primary motor area, however with a cluster of similarity of larger spatial extent in the postcentral gyrus adjacent to the superior parietal lobule (Fig. 4 (B), top panel), possibly involved in the sensation of the hand and fingers, and in an area of the postcentral gyrus adjacent to the supramarginal gyrus (Fig. 4 (B), bottom panel), involved in sensations of the face, mouth, or nose. Outside the postcentral gyrus, clusters found in the somatosensory translation that do not overlap with the motor translation are located in the central operculum (as defined in the Harvard-Oxford Cortical Structural atlas – Fig. 4 (B), bottom panel) and the parietal operculum, involved in somatosensory processing of laryngeal areas, as well as in the occipital pole.

The primary auditory cortex translation (Fig. 4 (C)) intersects with the somatosensory and motor translations in the posterior cingulate, the lateral occipital cortex, the occipital pole, middle frontal gyrus, and central operculum. It differs in the frontal pole, precuneus, angular gyrus, supramarginal gyrus, parietal operculum, superior frontal gyrus, and temporo-occipital middle temporal gyrus. Importantly, high similarity specific to the auditory translation is found in areas involved in primary and higher order auditory processing: the temporal pole, the planum polare, the anterior superior, middle and inferior temporal gyri, as well as in Heschl's gyrus, in the lingual gyrus, the temporal fusiform gyrus, the parahippocampal gyrus, and the posterior insula.

By looking at the differences between these primary translations, we highlight further patterns of similarity (Fig. 4 (D), rows 1-3). We observe that the primary somatosensory area shows greater similarity to primary and insular areas than the primary motor area, which is comparatively more similar to frontal, temporal, and parietal association areas (Fig. 4 (D), row 1). The auditory translation has a higher similarity to temporo-insular areas, areas typically involved in higher order auditory processing compared with the somatosensory translation (Fig. 4 (D), row 2), which highlights a parieto-frontal pattern of similarity, including areas more typically involved in higher order somatosensory processing.

Looking within modalities (Fig. 4 (D), rows 3, 4, 5), we can see that the auditory primary-association difference map (computed as the difference between temporal association and primary auditory cortex translations) is directed broadly from non-specific primary areas to anterior temporal, insular and medial frontal areas. The motor difference map is more diffuse across the cortex, from more focal primary areas – with emphasis on auditory – to parietal, fronto-temporal, and medial frontal areas. The somatosensory difference map, conversely, shows an emphasis on visual primary, and is directed more focally towards fronto-temporal and parietal areas.

On average (Fig. 4 (D), bottom row), we highlight a primary-association pattern of cross-species similarity, by creating a difference map between the average translation of mouse primary regions (motor, somatosensory, visual, auditory) and the average translation of mouse association regions to the human. On average, the mouse primary areas have higher similarity to human primary areas (in red in Fig. 4 (D), bottom row) than to human association areas, and mouse association areas have higher similarity to human association cortices (in blue in Figure 4 (D), bottom row) than human primary cortices. This is in line with previous results showing that a substantial portion of the variance in gene expression among cortical areas can be described as a gradient between primary areas and association cortex (Burt et al., 2018).

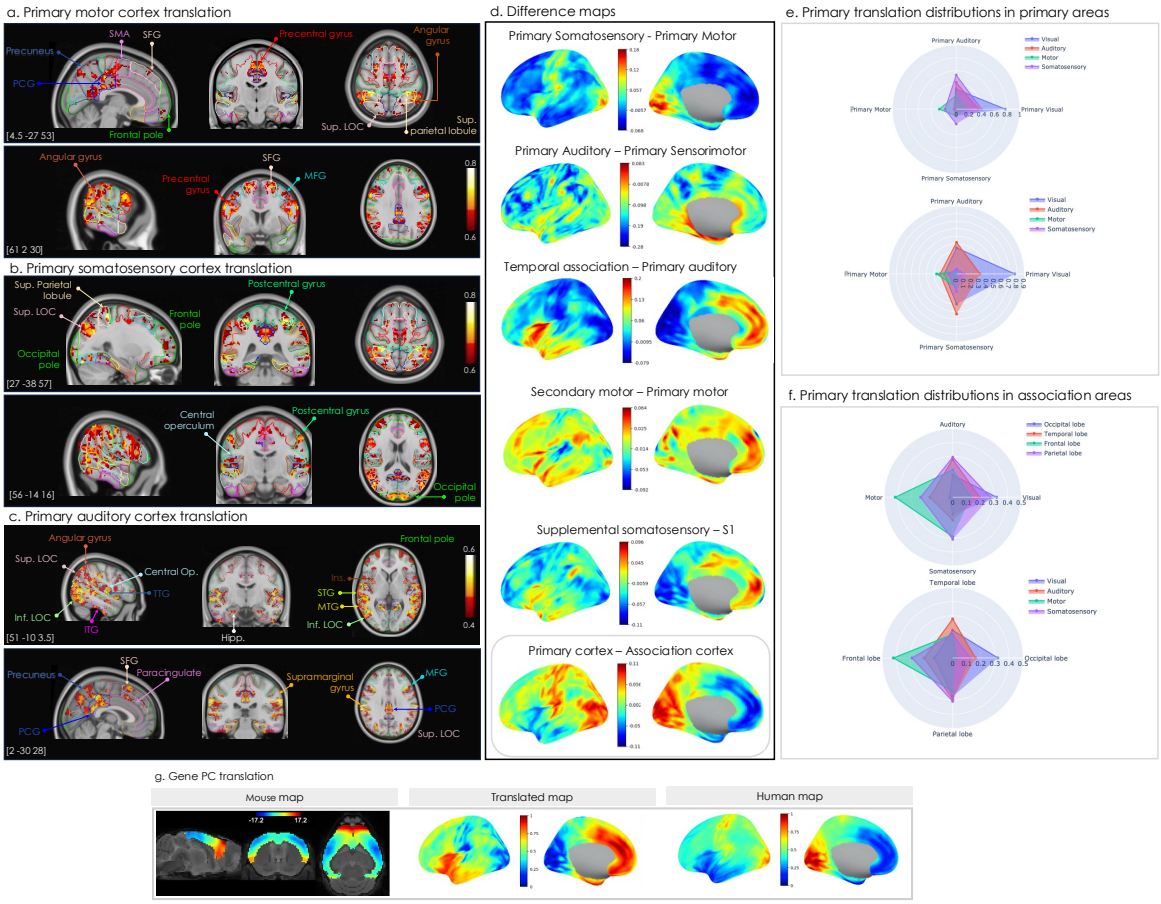

Fig. 2. Cross-species translations in neocortical regions. A, B, C – Thresholded voxel-wise volumetric translations from mouse primary motor, primary somatosensory, and auditory cortex. **D** – Surface representation of difference maps highlighting patterns of similarity within and across sensory modalities. **E** – Average distribution across human primary areas of translated voxels from mouse primary areas. The top row can read as 'Where does this mouse region land in the human brain?' and the second row can read as 'From which mouse area does this human area receive voxels?'. **F** – Average distribution across frontal, temporal, occipital and parietal lobes (association cortices) of the mouse primary translations. **G** – Translation of the gene expression PCA of the mouse and comparison with the gene expression PCA computed in the human.

To quantify the structure underlying these patterns of cortical similarity with respect to modality, we looked at the distribution of translated voxels from each primary modality within major anatomical areas (Fig. 4 (E, F)). First, looking at primary areas only (Fig. 4 (E)), we investigated the proportion of voxels within each primary area in the human cortex originating from each translation (top figure), and alternatively at the proportion of voxels in each primary translation landing within each primary region in the human cortex (bottom figure). Both cases confirm the

strength of the visual translation over other modalities, and similar patterns for the motor translation with smaller but consistent rates of voxels translating to target. The auditory and somatosensory translations show a degree of trade-off, depending on the angle of discussion. On one hand, we highlight that while the primary area receives more voxels from the somatosensory translation than from the auditory translation, the primary somatosensory area receives voxels mainly from the somatosensory translation. On the other hand, we highlight that the translation of the mouse auditory region goes first to auditory regions, and that the translation of the somatosensory area sends more voxels to the auditory and visual regions than to the somatosensory areas.

Looking at the distribution of primary translated voxels with respect to association areas, we highlight that primary translations predominantly remain within their corresponding association areas (Fig. 4 (F)). This is particularly clear for the visual and motor areas. The parietal lobe appears more multimodal, but still retains most voxels from the somatosensory modality. The auditory modality projects to both the parietal lobe and the temporal lobe, but we can see that the temporal lobe receives primarily voxels from the auditory regions. This confirms that there exists a within-modality pattern of cross-species similarity.

Cortical areas thus seem to translate from the mouse to the human in a reliable way. The fact that the human cortex is larger, more complex and subdivided could explain why single mouse regions tend to show high similarity to a series of human regions. This one-to-many mapping shows two overlapping principles of organization. First is a translation gradient between primary and association cortices (Fig. 4 (D) – bottom row); mouse primary areas are more similar to human primary areas, while mouse association areas are more similar to human association areas. Second is a within-modality gradient, whereby primary areas tend to show high similarity to areas of association cortex generally associated with processing the same type of information, as demonstrated in Fig. 3 (F) and illustrated in Fig. 4 (D) (rows 4 and 5).

2.3. Translating a cortex-wide scalar map

Moving beyond regional translations, we performed a first cortex-wide translation of a quantitative map (Fig. 4 (G)). Using the mouse-human similarity matrix, a scalar map in the mouse cortex was projected onto the human cortex by multiplying with the similarity matrix.

In humans, the first principal component of gene expression maps across brain-specific genes has been shown to correlate with the primary-association gradient of cortical hierarchy (Burt et al, 2018). We tested whether this principle was mirrored in the mouse brain, and if the translation of the first principal component of the mouse gene expression data matched the human gene expression PC1.

We ran the principal component analysis in both species on the subset of mouse-human ortholog genes used in our analysis. The second component highlighted a pattern consistent with a primary-association gradient, replicating findings from (28). For the human PCA, the first component was most consistent with a primary-association gradient and correlated significantly with the gene expression PC provided with the *Abagen* toolbox (22,29) (r=0.70, p<0.001). We thus performed the translation of the mouse second component, and since this translation was restricted to cortical areas only, we multiplied the mouse data by the cortical part of the latent similarity matrix. We compared the translated map with the human gene PC1, using *neuromaps*' statistical tools for spin tests accounting for spatial autocorrelation. The translated map correlated significantly with the human map (r=-0.9, p<0.001). We can note that the sign is negative because the mouse PCA situates primary cortices towards the lowest values while the human PCA does the opposite.

Having achieved a formal anatomical comparison of the mouse and human brain on the basis of the similarity of their embeddings in VAE latent space, we now move on to our second goal: translating disease phenotypes between species.

3. Translation of disease phenotypes across species

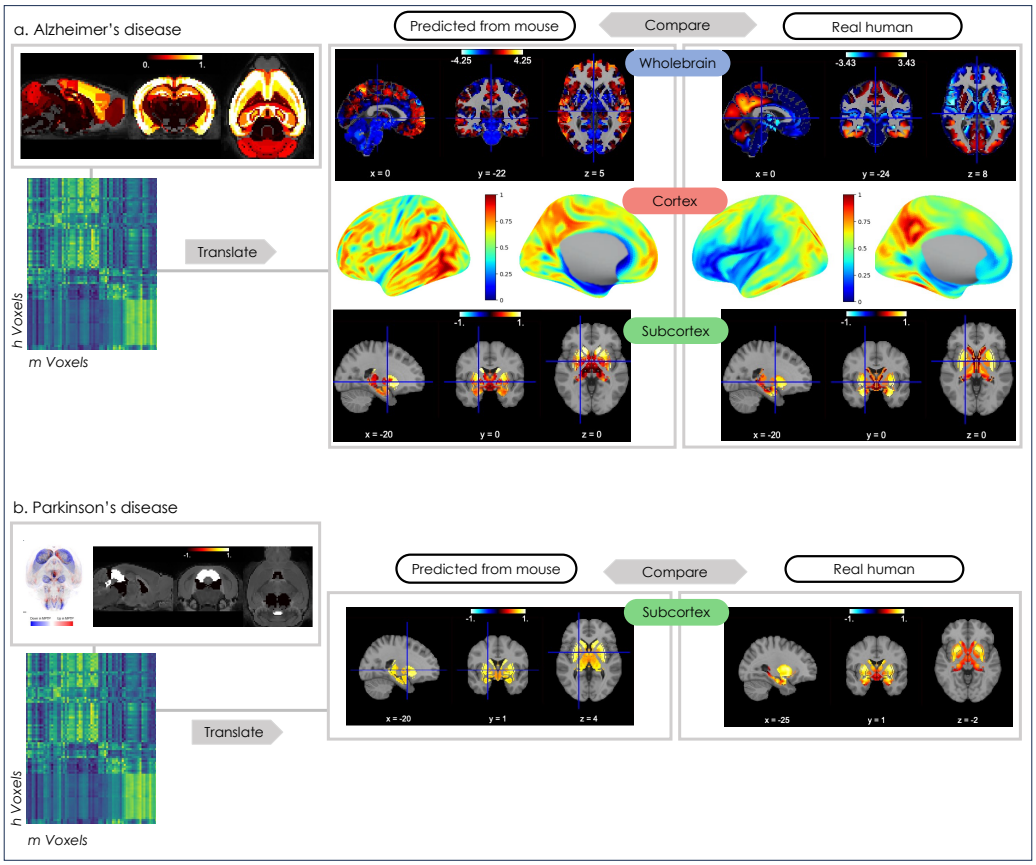

Fig. 5. A – Translation of an existing map of plaque deposition in a mouse model of Alzheimer's disease (APP-PSI, 13 months), divided between whole brain ($r=0.11$, $p<0.001$), cortical only ($r=0.22$, $p<0.001$), and subcortical only ($r=0.18$, $p<0.001$) translations. **B** – Subcortical translation of a mask of regions showing effects of disease based on an MPTP mouse model ($r=0.38$, $p=0.039$).

3.1. Translating a map of Alzheimer's disease

To demonstrate the use of the latent space to translate disease phenotypes, we performed whole-brain, cortical and subcortical translations of amyloid plaque density in the mouse AAP-PSI model at 13 months (Fig. 5 (A)), corresponding to an advanced stage of the disease (30). This model overexpresses mutant forms of the human amyloid precursor protein resulting in plaque formation due to amyloid beta deposition in the brain, first in the isocortex and with regional specificities. For comparison, we obtained a statistical map quantifying the effects of Alzheimer's disease compared to healthy controls in the human, from a study investigating networks of regions presenting grey matter volume decline, or synchronised degeneration networks (31). This study highlighted the posterior cingulate cortex and the hippocampus as vulnerable epicentre regions in Alzheimer's disease, and their degeneration networks as a marker of disease progression.

We first translated the map of amyloid deposition in the mouse and compared it the human brain atrophy map (Fig. 5 (A)). The correlation between the translated whole-brain map and the real whole-brain human map ($r=0.10$, $p<0.001$) was significant, but lower than the correlation when looking at the cortical data only ($r=0.23$, $p<0.001$). The subcortical map (containing subcortical areas and hippocampal subfields) correlated significantly with the real subcortical data ($r=0.18$, $p=0.008$).

From the whole brain comparison plots in Fig. 5 (A), we can see consistencies with the pattern of brain atrophy in the real data in the posterior cingulate cortex, the precuneus and occipital lobe, as well as in the temporal lobe. The main differences between the translation and the real data are located in the cerebellum and in the frontal areas. However in these areas, the translation remains consistent with the original mouse map, showing respectively low and high plaque deposit. The cortical surface representation further emphasises a pattern of lower atrophy in primary areas, both in the real human map and in the translated map, in addition to the high atrophy component in the posterior cingulate cortex. The subcortical plots highlight that despite not reaching significance, the pattern of cortical atrophy is well mirrored in parts of the caudate nucleus and putamen, as well as in the hippocampus highlighting a consistent gradient along the longitudinal axis, with increased atrophy in the head compared to the body and tail.

3.2. Translating a map of Parkinson's disease

Quantitative whole brain maps of a disease are not always readily available. Here, we wanted to give a different example of translation, where we leverage less information but still obtain a basis of comparison with an existing human quantitative map. We created a map consisting of subcortical regions where effects of disease were present, from a whole brain 3D light-sheet fluorenscence microscopy image quantifying tyroxine hydroxylase-positive neurons in MPTP mice compared to controls (32). In this widely used mouse model of Parkinson's disease, diseased mice exhibited loss of TH signal intensity in the substantia nigra, caudate nucleus, putamen, globus pallidus and subthalamic nucleus, and increased TH signal intensity mainly in the amygdala and hypothalamus. To recapitulate these results, we gave regions with loss of signal a coefficient of -1, and regions with increased signal a coefficient of +1 and translated this binary summary image.

We compared the translation to a statistical map of PD > Controls (Neurovault.collection:2694,(33)). In this study, functional connectivity differences with seeds in the basal ganglia were investigated between Parkinson's disease patients and healthy aging controls. Reduced functional connectivity with the basal ganglia network was found and this was improved through medication, highlighting the potential of the method to serve as a relevant biomarker for the disease. In our case, this statistical map provides crucial patterns of a metric with patterns of both increase and decrease due to disease, which is consistent with the translation we performed from the mouse.

Given that the mouse map consisted of subcortical regions, we performed a subcortical translation only, meaning that we took the sub-matrix of the similarity matrix corresponding to the subcortical voxels in both species. We found a significant correlation between the translated subcortical image and the real subcortical human data ($r=0.38$, $p=0.039$).

Discussion

Translational neuroscience suffers from a lack of formal comparative models. The present work takes a step towards predicting human outcomes based on preclinical findings. To interface between the mouse and the human, separated by over 85 million years of evolution, we built a common space based on the spatial patterns of expression of gene orthologs. Using this common space, we first demonstrated the ability of our approach to identify known cross-species homologies, before investigating how mouse regions translate across the much-expanded human neocortex, highlighting a one-to-one mapping of regions in all major anatomical subdivisions, as well as broader yet anatomically meaningful matches across regions within these subdivisions. Moving beyond anatomical comparisons, we exploited the common space to predict patterns of cortical atrophy in Alzheimer's disease based on a map of plaque deposition in the mouse brain (ref). Finally, we found a significant correlation between the translation of MPTP mouse map of disease effects and differences in functional connectivity in the human. Our model has thus shown

its ability to capture enough underlying information from the mouse gene expression data it was trained on, to generate a cross-species mapping tool capable of translating quantitative data across species.

Our common space took advantage of the availability of whole-brain transcriptomic maps in both the human and the mouse (12,13). Patterns of spatial gene expression have previously been used to show similarities in brain organization across the mouse and the human (34) and compare the effects of sex differences across species (35). The present work formalized this approach by in effect describing both the mouse and human brain in a common transcriptomic space, similar to what has previously been done for primate brains in connectivity space (19,36). However, the strong autocorrelation across genes and the sheer size number of data features mean that a data reduction is likely to improve the common space (23,37).

The present study uses a variational autoencoder to leverage the specific characteristics of its latent space. We trained the model on the mouse data, which is both the simplest of the two brains and the 'model' species based on which inference need to be made in translational neuroscience. Pragmatically, the coverage of the mouse transcriptomic data was also complete, while in the human dataset this was achieved through interpolation. We note that an alternative approach was taken by a recent study (38), but this required incorporation of additional receptor data as well as structural priors in a graph-based architecture. In our approach, the VAE itself provided us with a handle on the level of clustering versus scattering of the latent vectors, which is achieved through the weighting of the terms of the loss function. Furthermore, by combining the VAE with a latent classifier, we forced the latent space to remain sensitive to regional boundaries. We were thus able to achieve a balance between cross-species alignment and within-species regional clustering, in a semi data-driven way. As a result, our approach achieved significant improvement of the mouse human comparison compared to earlier approaches.

We provide the first systematic translation of binary regions across the entire mouse brain, yielding human voxel-wise distribution of values of similarity to each mouse region. We highlight the successful correspondence of most known homologs, with focal, quasi one-to-one matching across all major anatomical areas. While a complete one-to-one mapping would not be a faithful representation of the reality of cross species divergence, retrieving known homologs with precision is a reasonable expectation. Standing out among the most focal translations, our hippocampal results are consistent with recent studies of cross species alignment (37), and in-depth comparison of macaque and human hippocampus (39). This translation highlights increased similarity in the head of the hippocampus, but across all subfields, which is consistent with patterns of molecular variation along the longitudinal axis of the hippocampus, rather than the distal-proximal, as documented in the human (40).

We also highlight a number of one-to-many correspondences, in the cases where the translation spans beyond its supposed homologous area, with meaningful underlying patterns of brain organisation. Understanding such patterns is important, since they provide indications of the

potential success of using mouse models, even if the areas in the human brain have appeared after the separation of rodent and primate lineages, as was demonstrated in a recent study comparing the mouse and human premotor cortex (41). In our case, the beyond-target translations always remain within the main anatomical constituent tissue, and are mainly found in the cortex. They reveal two complementary patterns of brain organisation when looked at in combination. The first is a gradient between primary and association cortices, which is a well-known finding, replicated across modalities and species (28,34,42). The second is a modality-specific pattern whereby primary sensory areas have higher similarity to sensory association areas corresponding to the same modality.

Ultimately, the goal of our approach is to establish a quantitative footing for translational neuroscience, by enabling maps of brain organization or brain changes due to disease or treatment to be directly compared across species by translating through the common space. Here, we focused on two examples of neurodegenerative diseases, where translation rates from the preclinical setting to patient population are particularly low, and for which the data is heterogeneous across species. We demonstrate our model's ability to translate statistical results originating from different data modalities.

In humans, one aspect of Alzheimer's disease is amyloid plaque deposition, which has been shown to start in the cortex (43,44), progressing to the hippocampus, striatum, basal forebrain, thalamus, and to arise at the later stages in brainstem nuclei and cerebellum (45–47). Based on this spatial progression profile, the APP-PS1 model has been suggested to best mirror the later stages of the disease (30). We tested this observation through common space translation of the 2-photon imaging map of the mouse model, and found that it correlated significantly with human cortical and subcortical atrophy maps (31). This holds promise for future work involving extending the approach to simultaneous comparison of multiple mouse AD models at multiple timepoints, with different subtypes and stages of the disease in humans.

In Parkinson's disease, changes are first observed in the brainstem nuclei, with the loss of dopaminergic neurons in the substantia nigra, which in turn leads to compensation mechanisms in the basal ganglia and disruption of its networks. We were able to retrieve a similar result by translating a map of statistical differences between MPTP mice and wild type mice. The translation of disease effects in the mouse correlated significantly with changes in resting state functional connectivity in Parkinson's disease patients. This result holds promise in terms of being able to make inferences even when the best possible data is not readily available for translation.

The increasing availability of neuroimaging-derived phenotypes in both mice (48,49) and humans (50,51) means that the data are now available to scale our approach toward multi-level comparisons. While matching data modalities is often sought after in comparative studies, here we prove that we can manage heterogeneity and thus harness state of the art approaches for each species. Increasing data from mouse models and human disease will allow more direct comparisons, including elucidating where models fail, and which aspects of a disease are captured

by the variety of available models (10). Potentially more impactful even is the ability to link groups of patients in heterogenous syndromes, such as often seen in neurodevelopment and psychiatry, where clustering approaches have been able to identify subtypes both in humans (52) and mice (53). In these cases, mouse models of a disease often only capture a particular variant or aspect of its etiology; direct between-species comparisons have the potential to link models to specific subgroups of human patients.

In conclusion, we highlight a flexible and scalable framework to support translational neuroscience investigations ranging from fundamental enquiries to disease outcome prediction. Preclinical research is leaning towards more diverse animal models, while interest in personalized medicine is growing in the clinical setting. Our approach may eventually lead to answering the question: 'what is the best animal model for this patient?'

MATERIALS AND METHODS

Data and preprocessing

Mouse in-situ hybridisation data

The mouse data were taken from the AMBA Allen Institute database, and consisted of in situ hybridisation of an adult mouse brain (13), aligned to a 3D reference atlas, the Allen Mouse Brain Common Coordinate Framework (CCFv3) (54), with a resolution of 200µm. The majority of the analysis is based on the coronal dataset, but the sagittal dataset was used as well during the cross-validation of the model. Therefore, both the coronal and the sagittal datasets were downloaded using the Allen Institute's API, and saved as 3D Medical Image NetCDF (MINC) in RAS orientation. In the sagittal genes set, we kept the genes that belonged to the intersection with the coronal genes set.

Using the pyminc toolbox for analysing MINC files in python, the sagittal and coronal images were loaded, masked, reshaped, and concatenated into voxel-wise gene expression matrices, which included genes with multiple experiments. The coronal and sagittal expression matrices were log2 transformed, and the replicated experiments were averaged. Genes with more than 20% of voxels containing missing values were discarded. Below this threshold, missing values were imputed using K-nearest neighbours. Both gene sets were intersected with the set of genes common to human and mouse, resulting in a 2835 x 61315 (genes x voxels) coronal matrix, and a 2486 x 26317 (genes x voxels) sagittal matrix. Prior to further embedding or similarity computation, the data was normalized using z-score across all genes for each voxel, and de-meaned across voxels for each gene.

A region-wide matrix was generated by averaging values across regions, according to a modified version of the DSURQE atlas (55) to match the AMBA hierarchical ontology (23). White matter and ventricles were removed, and grey matter regions were aggregated to enable sufficient amounts of voxels per region for downstream training of deep learning algorithms. The resulting resolution of the atlas was of 67 grey matter regions. Annotations for 11 and 5 broad regions were also generated for visualisation purposes.

Resampled mouse data for independence during cross-validation

The following resampling of the mouse sagittal and coronal data was performed to maximise the independence of the data during the multiple training and validation rounds of the model for hyperparameter selection (23). First, the coronal data was masked using the sagittal mask, for consistent spatial extent. Then, for each gene, if multiple experiments were present in the coronal set, we randomly sampled one of those for the training set, and one for the validation set. If only one experiment existed for a given gene, then we randomly assigned the sagittal or the coronal to either training or validation.

Human microarray expression data

The human data were downloaded and pre-processed using the abagen python toolbox (version 0.1.3; https://github.com/rmarkello/abagen). To maximize spatial coverage, we generated dense gene expression maps from the full dataset provided by the Allen Human Brain Atlas (AHBA, https://human.brain-map.org; [H2012N]), consisting of regional microarray expression data from six post-mortem brains (1 female, ages 24.0–57.0, 42.50 +/- 13.38). Preprocessing followed the workflow of abagen's get_interpolated_map.

To generate interpolated maps, microarray probes were first reannotated, and probes missing a valid Entrez ID were discarded. The remaining probes were filtered based on their expression intensity relative to background noise (56): probes below background intensity in over 50% of samples across donors were also discarded. When multiple probes indexed the expression of the same gene, we selected and used the probe with the most consistent pattern of regional variation across donors (i.e., differential stability; (57)) calculated using:

$$\Delta_S(p) = \frac{1}{\binom{N}{2}} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \rho\left[B_i(p), B_j(p)\right]$$

Where ρ is Spearman's rank correlation of the expression of a single probe p, across regions in two donor brains B_i and B_j, and N is the total number of donors. For the purpose of this computation, regions are taken from the AHBA ontology.

The MNI coordinates of the tissue samples were updated from the original AHBA coordinates to those generated using Advanced Normalization Tools (ANTs) non-linear registration (https://github.com/chrisfilo/alleninf). Samples were mirrored across hemispheres to increase spatial coverage (lr_mirror set to bidirectional)

To address intersubject variation, tissue sample expression values were normalized across genes using a robust sigmoid function (58):

$$x_{norm} = \frac{1}{1 + e^{-\frac{x - <x>}{IQR_x}}}$$

where $< x >$ is the median and IQR_x is the normalized interquartile range of the expression of a single tissue sample across genes.

Normalized expression values were linearly rescaled to the unit interval:

$$x_{scaled} = \frac{x_{norm} - \min(x_{norm})}{\max(x_{norm}) - \min(x_{norm})}$$

Gene expression values were then normalized across tissue samples using the same procedure.

Interpolation was subsequently computed for each gene using K-nearest neighbour regression to estimate values at every voxel inside the provided mask. The resulting 1 mm resolution 3D gene expression maps with dimensions (197, 233, 189) were flattened and concatenated into a gene expression matrix spanning the 2835 common gene set, out of the total 15627 human genes. We

additionally normalized the data using z-score across all genes for each voxel, and subtracting the mean across voxels for each gene.

A region-wide matrix was generated by aggregating values per region, with voxel to region mapping performed using nilearn's NiftiLabelsMasker. For identification of brain areas, we used a modified version of the Allen Human Brain Atlas (27) initially based on the 2D version of the cellular resolution atlas of the adult human brain (59). We replaced the cerebellum in this atlas by the FSL cerebellar atlas, in which bilateral regions were merged for consistency with the rest of the atlas and the mouse data and removed white matter and ventricles.

Model architecture

We designed a model which consisted of an unsupervised variational autoencoder (VAE) (60) combined with a classifier (Fig.1 (A)). We implemented, trained, and applied the model in Python, using the Pytorch Lightning framework (61).

The VAE was tasked with reconstructing the mouse gene expression input based the latent representation. While deterministic autoencoders have for one input vector, one latent vector, and one corresponding decoded vector, here the VAE leveraged variational inference to encode the data onto multivariate gaussian distributions, from which the latent vectors were sampled for decoding. To avoid the issue of backpropagation bottleneck due to random sampling, the latent vector was expressed as a differentiable function:

$$z(x) = \mu(x) + \sigma(x) \cdot \epsilon$$

Where x is the input vector of the encoder, μ and σ are the output of the encoder, mean and standard deviation of the latent distribution from which we are sampling, and ϵ is a random variable sampled from the standard normal $N(0,1)$ independently of x. Simultaneously, in a supervised approach, the classifier was trained in the latent space of the variational autoencoder to recognize the 67 mouse atlas regions, thus taking as input the latent representation of the data instead of the high dimensional gene expression data.

The three modules of the model (encoder, decoder, classifier) are constructed as fully connected, multilayered networks. The input layer of the encoder and the output layer of the decoder had a fixed number of nodes, equal to the number of genes in the study. The output layer of the classifier was equal to the number of atlas regions. The input layer of the decoder and the input layer of the classifier had a number of nodes equal to the latent dimension, an adjustable hyperparameter. The output layer of the encoder is twice that dimension since it encodes both mean and standard deviation. The number of hidden layers of fully connected nodes with their non-linear activation functions was also adjusted for each module. The three modules were linked together in the loss function, which the model aimed to minimize at each iteration. The loss function consisted of two terms, to jointly train the variational autoencoder and the classifier. We used cross-entropy for the classification loss, with per-class rescaling weights to mitigate the issue of class imbalance during the training phase only:

$$L_{class}(x, y) = \frac{\sum_{n=1}^{N} l_n}{N}, \quad l_n = -w_{y_n} log \frac{\exp(x_{n,y_n})}{\sum_{c=1}^{C} \exp(x_{n,c})}$$

Where C is the number of classes, N is the batch dimension, x corresponds to the input logits (unnormalized prediction scores, before softmax application) of the classifier with dimensions (N, C), y is the target tensor with shape (N). In addition, w_{y_n} is the weight of the class y_n taken as the inverse of the frequency of the class.

The loss for the variational autoencoder, or Evidence Lower Bound (ELBO), consisted of another two terms: a reconstruction term, to compare the output of the decoder to the input of the encoder, and a regularization term which forces the latent distributions to match the shape of the standard Gaussian distribution. We used mean squared error for the reconstruction loss and KL divergence for comparing the latent distribution to the standard Gaussian $\mathcal{N}(0, I)$:

$$L_{ELBO} = L_{rec.} + L_{reg.} = \frac{1}{2}\|x - \hat{x}\|^2 + \frac{1}{2}\sum_{i=1}^{d}(\sigma_i^2 + \mu_i^2 - 1 - log(\sigma_i^2))$$

Where x is the input vector, $\hat{x}$ is the reconstructed vector, d is the dimension of the latent space, σ_i and μ_i are the mean and standard deviations of the latent gaussian distribution over the latent space for input x . The total loss for the model was the sum of the classifier loss and the VAE loss.

Model training and optimisation

The model was trained and validated as described below. As it was no a priori known which hyperparameters (e.g., number of layers, number of nodes) were optimal, they were first "tuned", meaning that training and validation was performed for different combinations of hyperparameters, using the resampled dataset dedicated for this purpose. Then, the model with the best performance as measured in terms of accuracy in the classification task was trained and validated using the full dataset.

Training and Validation

The full mouse gene expression data was divided into 64 batches for the model to run over during the course of an epoch. At the end of an epoch, the loss was computed, the weights were adjusted accordingly, and another epoch was started. Via this iterative training process, the model converged towards a minimum of the loss function. To assess the progression of the performance of the model, we ran a validation step at the end of each epoch. The dataset was separated into training and validation sets, using two approaches to maximise independence. During the hyperparameter search, we used the dedicated resampled mouse data, whereby the random assignment and shuffling of coronal and sagittal data ensured a different set of data for each run. For the model chosen through the automated hyperparameter search, we randomly assigned 80% of the full coronal dataset to training, and 20% to validation.

During the hyperparameter search, each model was trained over 50 epochs, while the final model was trained over 150 epochs.

Hyperparameter selection

A tuning process was conducted to empirically decide on the type of activation function (ReLU, SiLU), the type of normalization (batch or instance), the number of hidden layers (2 or 3) and the number of nodes per layer (100, 150, or 200), the latent dimension (10,50,100,200), the amount of dropout (0, 0.1, 0.2, 0.4), the weight of the classifier loss (1, 0.5), the reconstruction weight (1, 0.5), the KL weight (0.001, 0.002, 0.005), and the maximum learning rate (10^{-5}, 10^{-4}, 10^{-3}). To optimize this process, we used Optuna, a python based hyperparameter tuning algorithm, which sampled a subset of hyperparameters from all possible combinations, using a Bayesian optimization technique *(TPEsampler*), and trained the corresponding versions of the model over 50 epochs maximum with an early stopping callback to stop training after 3 epochs with no improvement. Model performance was measured using the multi-class macro-average accuracy score in the classification task, chosen to mitigate class imbalance.

The optimisation process highlighted that the best performing model had a latent dimension of 50 nodes, an encoder with two hidden layers and 200 nodes per layer, a decoder with two hidden layers and 100 nodes per layer, a classifier with three hidden layers and 150 nodes per layer, a SiLU activation function, instance normalisation, a 10% dropout, a classifier and reconstruction weight of 1, a kl coefficient of 0.001, and a max learning rate of 10^{-5}. We optimized the model using AdamW (62). The maximum training accuracy over the 50 epochs was 0.865, and the maximum validation accuracy was 0.565.

Comparison with previous models

We compared the performance of the cross-species comparison using our model with the cross-species comparison obtained using the gene expression data with no dimensionality reduction, and using the multi-layer perceptron developed in previous work (23).

First, we performed a 2-dimensional UMAP embedding (24) to visually highlight changes in the distribution of the data across models. Then, we computed the gene expression cross-species similarity and the MLP-embedding cross-species similarity, to highlight how the distribution of similarity values on the heatmap change with regards to the model. We performed the hierarchical clustering on all three matrices, whereby smaller clusters indicate that one region in the source species has fewer high similarity matches, i.e. potential homologs, in the other species. This highlights changes in the focality of the comparison across the different models. To investigate differences in broad anatomical patterns, we added labels to the hierarchical clustering for primary, association, and limbic cortex, subcortex, hippocampus, cerebellum, and brainstem. We quantified the differences in the regional similarity matrices by performing a Kolmogorov-Smirnov test.

In addition, we investigated differences in unit-scaled rank-ordered similarity profiles, across all regions taken together, and for each mouse region separately. First, this resulted in a vector of 67 x 129 regions. Secondly, this resulted in 67 profiles, one for each mouse region, each with a length of 129, corresponding to the number of human regions as defined by the atlas used in this study.

A steeper decay rate of the ordered similarity profile highlights a more focal comparison. Differences in rank ordered profiles were tested using the two-sample Kolmogorov-Smirnov test.

Model application

Data embedding and similarity

Once the optimal model was trained, we used the encoder module to project data from gene expression space to latent space, by designing the model to return the output of the encoder when calling *model.predict* in Pytorch Lightning. We transformed both the human and mouse voxel-wise and region-wise expression matrices into the latent space. Using the latent sampled vectors, we computed the similarity between the latent matrices using Pearson's correlation, at the regional and the voxel level.

Cross-species translation of anatomical areas

We used the voxel similarity matrices to investigate the distribution of cross-species similarity values across the whole human brain and the whole mouse brain. To optimize computation time, the voxel matrices mapped of only one human hemisphere. Therefore, each column in the matrix mapped one human hemisphere, and each row in the matrix mapped a full mouse brain. We define as translation the dot product between a vector mapping a distribution of values in one species and the similarity matrix. In the mouse to human direction, we reconstructed the brains by mirroring the data to map two hemispheres for visualisation purposes. We visualised human volumetric brain maps using FSLeyes as well as nilearn's plotting tools (nilearn.plotting.plot_stat_map and view_img). We provided surface-based representations using nilearn's surface package (nilearn.surface). For the cerebellum (Supplementary Table S1), we visualised the translations via flattened representations, using SUITpy.flatmap (25).

First, we investigated translations using a seed-based approach, to investigate the improvement in homology of the comparison. For 35 pairs of homologous regions based on previous work and literature, we multiplied the flattened binary mask image of the ROI by the similarity matrix, equivalent to taking the sum over the columns mapping that region. This yielded a whole human brain map of values ranging from least similar to most similar to the mouse seed region. We normalised the values between 0 and 1 using min-max scaling. We thresholded these maps to optimise the translation profile. Areas with narrow similarity profiles were given high thresholds, while areas with broader profiles were given lower thresholds, starting at 0.6. This threshold was adjusted to maximize the ratio of the number of voxels hitting the target ROI over the number of voxels in that target ROI. This was implemented in *python* using the *scipy.optimize.minimize* package. The exact threshold used for each area is documented in Supplementary Table S2.

We quantified how well each translation was hitting its target by defining three types of scores. We defined for each target area the rate of 'hits' as the number of voxels above threshold in the target divided by the total number of voxels in the target. We defined the rate of 'near hits' as the number of voxels translating outside the target but within the same constituent part of the brain as the target area (respectively cortex, subcortex, brainstem, or cerebellum), divided by the total number of voxels in a mask of the constituent area minus the target area. We defined a rate of

‘misses’ as the number of translated voxels falling in the rest of the brain, divided by the total number of voxels in the corresponding mask. We performed an ANOVA to check for an interaction between score type and constituent part of the brain, and Connover’s posthoc test to see between which pairs of constituent brain areas the interaction was.

Then, focusing on cortical areas, we investigated gradients of cross species similarity between brain areas by computing the difference between sets of translations. To check for a primary-association gradient, we looked at the difference between the average translation of all cortical association regions (Taenia tecta, cortical amygdalar area, piriform-amygdalar area, subiculum, entorhinal area, dorsal auditory area, ventral auditory area, secondary motor area, supplemental somatosensory area, posterior parietal association areas, temporal association areas, frontal pole, and orbital area) and primary cortex areas (primary motor, primary auditory, primary somatosensory, visual areas, main olfactory bulb). We also looked at within-modality differences, i.e primary motor versus secondary motor areas, primary auditory versus dorsal and ventral auditory, primary somatosensory versus supplemental somatosensory.

To further illustrate the modality-specific effect, we computed the proportion of each primary translation falling in its corresponding target area and in each other primary cortical area, as well as in each corresponding association cortex area. For the visual cortex, the occipital lobe served as proxy for the corresponding association cortex. Similarly, we used the temporal lobe for the auditory cortex, frontal lobe for motor cortex, and parietal lobe for somatosensory cortex.

Cross-species translation of gene expression PCA

We performed a principal component analysis using the mouse and human ortholog gene expression data of this study, using scikit-learn's PCA decomposition module. We masked the data to perform the PCA on the cortex only, therefore we also masked out non-cortical areas in the latent similarity matrix. We translated the first two components of the mouse and compared them to the first two components computed directly in the human. To test for statistical significance, we transformed the maps onto the fsaverage cortical surface and used spin tests in python using the neuromaps package (29) accounting for the spatial autocorrelation of the surface-based data. The nulls were generated in surface space using the *alexander_bloch* method.

Cross-species translation of Alzheimer’s disease map

We based our first disease phenotype translation on the 13-month-old PSI-APP mouse model of Alzheimer’s disease (30). We first coregistered the image to the template (DSURQE in CCFV3 space) used in this study. Then, we computed the cortical translation of this image, and the subcortical translation. We split the analysis to be able to run the cortical comparison on the fsaverage surface, regarded as the optimal solution in neuromaps, and the subcortical comparison in volume space.

We compared the translated map to the cortical atrophy map taken from (Lee et al, 2020), neurovault.collections:3273. We coregistered this map to the human atlas used in this study using

neuromaps' *transforms* module from MNI to MNI space to transform between resolutions and symmetrised the data for consistency throughout our study. For the cortical data, we computed Pearson's correlation and tested for significance using *neuromaps.stats* and the *alexander_bloch* method. For the subcortical data, we used Pearson's correlation and tested for significance using the *burt2018* model for generating spatial nulls in volume space.

Cross-species translation of Parkinson's disease map

To translate results from mouse to human in Parkinson's Disease, we used the results of previous study to constitute a mask of effects in the mouse (32). We attributed a (+1) value for regions where MPTP mice had increased tyrosine-hydroxylase expression compared to controls, and (-1) value for regions where this MPTP mice had lower TH expression compared to controls. MPTP mice had reduced TH expression in basal ganglia nuclei including midbrain areas substantia nigra and substantia nigra pars compacta, caudoputamen, pallidum, subthalamic nucleus. There was an increased TH expression mainly in limbic regions including regions of the amygdala and hypothalamus.

We computed the translation of this multi-region mask and compared it to the human statistical map taken from neurovault.collections:2694 (33), choosing the contrast PD>NC. We computed Pearson's correlation and tested for significance in volume space using *burt2018* nulls in *neuromaps.stats.compare_images*.

References

1. Rigdon G. 1192U90 in Animal Tests That Predict Antipsychotic Efficacy, Anxiolysis, and Extrapyramidal Side Effects. Neuropsychopharmacology. 1996 Sep;15(3):231–42.

2. Naryshkin NA, Weetall M, Dakka A, Narasimhan J, Zhao X, Feng Z, et al. *SMN2* splicing modifiers improve motor function and longevity in mice with spinal muscular atrophy. Science. 2014 Aug 8;345(6197):688–93.

3. Steinman L, Zamvil SS. How to successfully apply animal studies in experimental allergic encephalomyelitis to research on multiple sclerosis. Ann Neurol. 2006 Jul;60(1):12–21.

4. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, et al. Multiplex Genome Engineering Using CRISPR/Cas Systems. Science. 2013 Feb 15;339(6121):819–23.

5. Hsu PD, Lander ES, Zhang F. Development and Applications of CRISPR-Cas9 for Genome Engineering. Cell. 2014 Jun;157(6):1262–78.

6. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. Science. 2012 Aug 17;337(6096):816–21.

7. Shultz LD, Ishikawa F, Greiner DL. Humanized mice in translational biomedical research. Nat Rev Immunol. 2007 Feb;7(2):118–30.

8. Walsh NC, Kenney LL, Jangalwe S, Aryee KE, Greiner DL, Brehm MA, et al. Humanized Mouse Models of Clinical Disease. Annu Rev Pathol Mech Dis. 2017 Jan 24;12(1):187–215.

9. Hay M, Thomas DW, Craighead JL, Economides C, Rosenthal J. Clinical development success rates for investigational drugs. Nat Biotechnol. 2014 Jan;32(1):40–51.

10. Striedter G. Model Systems in Biology: History, Philosophy, and Practical Concerns [Internet]. The MIT Press; 2022 [cited 2024 Dec 11]. Available from: https://direct.mit.edu/books/book/5357/Model-Systems-in-BiologyHistory-Philosophy-and

11. Kumar S, Stecher G, Suleski M, Hedges SB. TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. Mol Biol Evol. 2017 Jul;34(7):1812–9.

12. Hawrylycz MJ, Lein ES, Guillozet-Bongaarts AL, Shen EH, Ng L, Miller JA, et al. An anatomically comprehensive atlas of the adult human brain transcriptome. Nature. 2012 Sep;489(7416):391–9.

13. Lein ES, Hawrylycz MJ, Ao N, Ayres M, Bensinger A, Bernard A, et al. Genome-wide atlas of gene expression in the adult mouse brain. Nature. 2007 Jan;445(7124):168–76.

14. Ortiz C, Navarro JF, Jurek A, Märtin A, Lundeberg J, Meletis K. Molecular atlas of the adult mouse brain. Sci Adv. 2020 Jun 26;6(26):eabb3446.

15. Yao Z, Van Velthoven CTJ, Kunst M, Zhang M, McMillen D, Lee C, et al. A high-resolution transcriptomic and spatial atlas of cell types in the whole mouse brain. Nature. 2023 Dec 14;624(7991):317–32.

16. Amunts K, Zilles K. Architectonic Mapping of the Human Brain beyond Brodmann. Neuron. 2015 Dec;88(6):1086–107.

17. Paxinos G, Franklin KBJ. Paxinos and Franklin's The mouse brain in stereotaxic coordinates. Fifth edition. London: Academic Press, an imprint of Elsevier; 2019. 1 p.

18. Mars RB, Jbabdi S, Rushworth MFS. A Common Space Approach to Comparative Neuroscience. Annu Rev Neurosci. 2021 Jul 8;44(1):69–86.

19. Mars RB, Passingham RE, Jbabdi S. Connectivity Fingerprints: From Areal Descriptions to Abstract Spaces. Trends Cogn Sci. 2018 Nov;22(11):1026–37.

20. Warrington S, Thompson E, Bastiani M, Dubois J, Baxter L, Slater R, et al. Concurrent mapping of brain ontogeny and phylogeny within a common space: Standardized tractography and applications. Sci Adv. 2022 Oct 21;8(42):eabq2022.

21. Balsters JH, Zerbi V, Sallet J, Wenderoth N, Mars RB. Primate homologs of mouse cortico-striatal circuits. eLife. 2020 Apr 16;9:e53680.

22. Burt JB, Demirtaş M, Eckner WJ, Navejar NM, Ji JL, Martin WJ, et al. Hierarchy of transcriptomic specialization across human cortex captured by structural neuroimaging topography. Nat Neurosci. 2018 Sep;21(9):1251–9.

23. Beauchamp A, Yee Y, Darwin BC, Raznahan A, Mars RB, Lerch JP. Whole-brain comparison of rodent and human brains using spatial transcriptomics. eLife. 2022 Nov 7;11:e79418.

24. McInnes L, Healy J, Melville J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction [Internet]. arXiv; 2018 [cited 2025 Feb 28]. Available from: https://arxiv.org/abs/1802.03426

25. Nettekoven C, Zhi D, Shahshahani L, Pinho AL, Saadon-Grosman N, Buckner RL, et al. A hierarchical atlas of the human cerebellum for functional precision mapping. Nat Commun. 2024 Sep 27;15(1):8376.

26. Krubitzer L. The organization of neocortex in mammals: are species differences really so different? Trends Neurosci. 1995 Sep;18(9):408–17.

27. Ding SL, Royall JJ, Sunkin SM, Facer BAC, Lesnar P, Bernard A, et al. Allen Human Reference Atlas – 3D, 2020. http://download.alleninstitute.org/informatics-archive/allen_human_reference_atlas_3d_2020/version_1/; 2020.

28. Huntenburg JM, Yeow LY, Mandino F, Grandjean J. Gradients of functional connectivity in the mouse cortex reflect neocortical evolution. NeuroImage. 2021 Jan;225:117528.

29. Markello RD, Hansen JY, Liu ZQ, Bazinet V, Shafiei G, Suárez LE, et al. neuromaps: structural and functional interpretation of brain maps. Nat Methods. 2022 Nov;19(11):1472–9.

30. Whitesell JD, Buckley AR, Knox JE, Kuan L, Graddis N, Pelos A, et al. Whole brain imaging reveals distinct spatial patterns of amyloid beta deposition in three mouse models of Alzheimer's disease. J Comp Neurol. 2019 Sep;527(13):2122–45.

31. Lee PL, Chou KH, Chung CP, Lai TH, Zhou JH, Wang PN, et al. Posterior Cingulate Cortex Network Predicts Alzheimer's Disease Progression. Front Aging Neurosci. 2020;12:608667.

32. Roostalu U, Salinas CBG, Thorbek DD, Skytte JL, Fabricius K, Barkholt P, et al. Quantitative whole-brain 3D imaging of tyrosine hydroxylase-labeled neuron architecture in the mouse MPTP model of Parkinson's disease. Dis Model Mech. 2019 Nov 1;12(11):dmm042200.

33. Szewczyk-Krolikowski K, Menke RAL, Rolinski M, Duff E, Salimi-Khorshidi G, Filippini N, et al. Functional connectivity in the basal ganglia network differentiates PD patients from controls. Neurology. 2014 Jul 15;83(3):208–14.

34. Fulcher BD, Murray JD, Zerbi V, Wang XJ. Multimodal gradients across mouse cortex. Proc Natl Acad Sci. 2019 Mar 5;116(10):4689–95.

35. Guma E, Beauchamp A, Liu S, Levitis E, Ellegood J, Pham L, et al. Comparative neuroimaging of sex differences in human and mouse brain anatomy [Internet]. Neuroscience; 2023 Aug [cited 2024 Jan 25]. Available from: http://biorxiv.org/lookup/doi/10.1101/2023.08.23.554334

36. Bryant KL, Camilleri J, Warrington S, Freches GB, Sotiropoulos SN, Jbabdi S, et al. Connectivity profile and function of uniquely human cortical areas [Internet]. Neuroscience; 2024 [cited 2025 Feb 24]. Available from: http://biorxiv.org/lookup/doi/10.1101/2024.06.20.599486

37. Zhang B, Zhang S, Zhang S. Whole brain alignment of spatial transcriptomics between humans and mice with BrainAlign. Nat Commun. 2024 Jul 30;15(1):6302.

38. Huang S, Zhang T, Dong C, Shi Y, Peng Y, Liu X, et al. TransBrain: A computational framework for translating brain-wide phenotypes between humans and mice [Internet]. Neuroscience; 2025 [cited 2025 Feb 24]. Available from: http://biorxiv.org/lookup/doi/10.1101/2025.01.27.635016

39. Eichert N, DeKraker J, Howard AFD, Huszar IN, Zhu S, Sallet J, et al. Hippocampal connectivity patterns echo macroscale cortical evolution in the primate brain. Nat Commun. 2024 Jul 16;15(1):5963.

40. Vogel JW, La Joie R, Grothe MJ, Diaz-Papkovich A, Doyle A, Vachon-Presseau E, et al. A molecular gradient along the longitudinal axis of the human hippocampus informs large-scale behavioral systems. Nat Commun. 2020 Feb 19;11(1):960.

41. Lazari A, Tachrount M, Valverde JM, Papp D, Beauchamp A, McCarthy P, et al. The mouse motor system contains multiple premotor areas and partially follows human organizational principles. Cell Rep. 2024 May;43(5):114191.

42. Froudist-Walsh S, Xu T, Niu M, Rapan L, Zhao L, Margulies DS, et al. Gradients of neurotransmitter receptor expression in the macaque cortex. Nat Neurosci. 2023 Jul;26(7):1281–94.

43. Braak H, Braak E. Neuropathological stageing of Alzheimer-related changes. Acta Neuropathol (Berl). 1991 Sep;82(4):239–59.

44. Palmqvist S, Schöll M, Strandberg O, Mattsson N, Stomrud E, Zetterberg H, et al. Earliest accumulation of β-amyloid occurs within the default-mode network and concurrently affects brain connectivity. Nat Commun. 2017 Oct 31;8(1):1214.

45. Buckner RL, Snyder AZ, Shannon BJ, LaRossa G, Sachs R, Fotenos AF, et al. Molecular, Structural, and Functional Characterization of Alzheimer's Disease: Evidence for a Relationship between Default Activity, Amyloid, and Memory. J Neurosci. 2005 Aug 24;25(34):7709–17.

46. Grothe MJ, Barthel H, Sepulcre J, Dyrba M, Sabri O, Teipel SJ, et al. In vivo staging of regional amyloid deposition. Neurology. 2017 Nov 14;89(20):2031–8.

47. Rice L, Bisdas S. The diagnostic value of FDG and amyloid PET in Alzheimer's disease—A systematic review. Eur J Radiol. 2017 Sep;94:16–24.

48. Lerch JP, Sled JG, Henkelman RM. MRI Phenotyping of Genetically Altered Mice. In: Modo M, Bulte JWM, editors. Magnetic Resonance Neuroimaging [Internet]. Totowa, NJ: Humana Press; 2011 [cited 2025 Feb 24]. p. 349–61. (Methods in Molecular Biology; vol. 711). Available from: https://link.springer.com/10.1007/978-1-61737-992-5_17

49. Nieman BJ, Bock NA, Bishop J, Chen XJ, Sled JG, Rossant J, et al. Magnetic resonance imaging for detection and analysis of mouse phenotypes. NMR Biomed. 2005 Nov;18(7):447–68.

50. Soheili-Nezhad S, Schijven D, Mars RB, Fisher SE, Francks C. Distinct impact modes of polygenic disposition to dyslexia in the adult brain [Internet]. Genetic and Genomic Medicine; 2024 [cited 2025 Feb 24]. Available from: http://medrxiv.org/lookup/doi/10.1101/2024.02.14.24302815

51. Thompson PM, Jahanshad N, Ching CRK, Salminen LE, Thomopoulos SI, Bright J, et al. ENIGMA and global neuroscience: A decade of large-scale studies of the brain in health and disease across more than 40 countries. Transl Psychiatry. 2020 Mar 20;10(1):100.

52. Hrdlicka M, Dudova I, Beranova I, Lisy J, Belsan T, Neuwirth J, et al. Subtypes of autism by cluster analysis based on structural MRI data. Eur Child Adolesc Psychiatry. 2005 Jun;14(3):138–44.

53. Ellegood J, Anagnostou E, Babineau BA, Crawley JN, Lin L, Genestine M, et al. Clustering autism: using neuroanatomical differences in 26 mouse models to gain insight into the heterogeneity. Mol Psychiatry. 2015 Feb;20(1):118–25.

54. Wang Q, Ding SL, Li Y, Royall J, Feng D, Lesnar P, et al. The Allen Mouse Brain Common Coordinate Framework: A 3D Reference Atlas. Cell. 2020 May;181(4):936-953.e20.

55. Dorr AE, Lerch JP, Spring S, Kabani N, Henkelman RM. High resolution three-dimensional brain atlas using an average magnetic resonance image of 40 adult C57Bl/6J mice. NeuroImage. 2008 Aug;42(1):60–9.

56. Quackenbush J. Microarray data normalization and transformation. Nat Genet. 2002 Dec;32(S4):496–501.

57. Hawrylycz M, Miller JA, Menon V, Feng D, Dolbeare T, Guillozet-Bongaarts AL, et al. Canonical genetic signatures of the adult human brain. Nat Neurosci. 2015 Dec;18(12):1832–44.

58. Fulcher BD, Little MA, Jones NS. Highly comparative time-series analysis: the empirical structure of time series and their methods. J R Soc Interface. 2013 Jun 6;10(83):20130048.

59. Ding S, Royall JJ, Sunkin SM, Ng L, Facer BAC, Lesnar P, et al. Comprehensive cellular-resolution atlas of the adult human brain. J Comp Neurol. 2016 Nov;524(16):3127–481.

60. Kingma DP, Welling M. An Introduction to Variational Autoencoders. 2019 [cited 2025 Feb 28]; Available from: https://arxiv.org/abs/1906.02691

61. Falcon W, Borovec J, Wälchli A, Eggert N, Schock J, Jordan J, et al. PyTorchLightning/pytorch-lightning: 0.7.6 release [Internet]. Zenodo; 2020 [cited 2025 Feb 28]. Available from: https://zenodo.org/record/3828935

62. Loshchilov I, Hutter F. Decoupled Weight Decay Regularization. 2017 [cited 2024 Jan 25]; Available from: https://arxiv.org/abs/1711.05101

Acknowledgments:

Funding: This study was supported by:

- the BBSRC grant Computational comparative anatomy: Translating between species in neuroscience [BB/X013227/1],
- the MRC Fellowship Quantitative translational neuroscience: Bridging preclinical and human neuroscience research [MR/Y010698/1],
- the ERC Consolidator award (ERC 101000969 supporting MA and SNS)
- the NIHR Oxford Health Biomedical Research Centre (NIHR203316) with the Wellcome Centre for Integrative Neuroimaging supported by core funding from Wellcome [201319/Z/16/Z; 203139/A/16/Z].

The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care. For the purpose of open access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

Author contributions:

Conceptualization: CJ, SNS, RBM
Methodology: CJ, MA, AB, SNS, RBM
Funding acquisition: RBM, JPL, SNS
Supervision: RBM, JPL, SNS
Writing – original draft: CJ, RBM
Writing – review & editing: CJ, MA, AB, JPL, SNS, RBM

Competing interests: Authors declare that they have no competing interests.

Data and materials availability: Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Chloe Jaroszynski (chloe.jaroszynski@ndcn.ox.ac.uk).

Data and code availability

- Mouse and human transcriptomic data are openly available through the Allen Institute data portal https://portal.brain-map.org, and can be freely downloaded using the Allen Institute API (http://api.brain-map.org), or the abagen toolbox (available from: 10.5281/zenodo.3451463).
- The python and R code used for downloading and preprocessing the data, training the model, running the analysis and generating the figures is available on the Gitlab server of the Wellcome Center of Integrative Neuroimaging at https://git.fmrib.ox.ac.uk/chloe.jaroszynski/dl_mouse_human, as well as at https://github.com/ChloeJaro/dl_mouse_human.
- Any additional information required to replicate this study is available upon request to the lead contact.

Key resources table

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
3D expression grid AMBA data	Allen Institute data portal, (13)	https://mouse.brain-map.org
CCFv3 Coordinate Framework	(54)	https://atlas.brain-map.org/
DSURQE Atlas	(55)	(23) https://github.com/ChloeJaro/dl_mouse_human
AHBA data	Allen Institute data portal	https://human.brain-map.org
3D Human brain atlas	Adapted from	(27)
Software and Algorithms		

abagen	doi:10.7554/eLife.72129	https://github.com/rmarkello/abagen.git
Pytorch-lightning	doi: 10.5281/zenodo.3828935	https://lightning.ai
captum	https://doi.org/10.48550/arXiv.2009.07896	https://captum.ai
optuna	https://doi.org/10.48550/arXiv.1907.10902	Optuna.org
Processing code	This article	Github/ChloeJaro/dl_mouse_human