

# Transcriptomic similarity in the mouse and human brain

Antoine Beauchamp

December 03, 2021

(Make front matter)

## Abstract

(Write abstract)

## Introduction

Animal models play an indispensable role in neuroscience research, both in the context of preclinical translation and in that of basic science. While numerous species have been used to model the human brain, the mouse has emerged as the most prominent of these, due to its rapid life cycle, straightforward husbandry, and amenability to genetic engineering (Hedrich, Mossman, and Nicklas 2004; Houdebine 2004; Dietrich, Ankeny, and Chen 2014; Ellenbroek and Youn 2016). Mouse models have proven to be extremely useful for understanding diverse features of the brain, from its molecular neurobiological properties to its large-scale network properties. While studying the mouse brain is a worthwhile endeavour in its own right, the primary purpose of using the mouse as a model system is to make inferences about the human brain. Consequently it is crucial that neuroscientific findings discovered using mouse models be translatable to humans. In order to accomplish this translation, a correspondence must be established between the brains of the two species. This is not trivial, given that the mouse and human are separated by approximately 80 million years of divergent evolution. Still, their brains (and all mammalian brains) exhibit a number of architectonic and functional similarities at present day. These features, which are assumed to have been inherited from the last common ancestor, can be used to identify neuroanatomical regions that are deemed homologous (Kaas 2005). Examples of such features include cytoarchitecture, myeloarchitecture, inter-regional connectivity, response to sensory input, and neurodevelopmental origin (Kaas 1983; Rosa and Krubitzer 1999; Bronchti et al. 2002; Krubitzer and Kaas 2005). This approach has worked well to identify general correspondences between mouse and human brains, particularly in the subcortex and in primary isocortical areas, but it is not without limitations. Establishing neuroanatomical homologues in this way is an exercise that is primarily qualitative in nature: Regional definitions from one species, e.g. Brodmann's areas (Zilles 2018), are matched to regional definitions in the other species. This qualitative approach makes it difficult to assess the degree of similarity between regions that are thought to be homologous. Additionally, such semantic matches might obscure nuances in the ways that the brains of different species relate to one another. While the mammalian brain follows a basic organizational blueprint, the brains of different species have evolved in unique ways to

adapt to their specific ecological niches. These adaptations often lead to re-organization of brain regions in diverse ways that are not easily captured by one-to-one semantic matches (Krubitzer and Kaas 2005). A striking example is the blind mole rat, whose brain features an occipital area that can be identified as the primary visual area on the basis of architectonic analyses. Interestingly, a significant portion of this region receives auditory input from the inferior colliculus (relayed via the dorsal lateral geniculate nucleus) and is thus activated by auditory stimuli (Bronchti et al. 2002). A specialized region such as this one would be better described by a many-to-one or many-to-many relationship compared with the mouse or human brain. These limitations highlight the need for novel rigorous ways of comparing brains across species.

Over the last decade, researchers in the field of comparative neuroanatomy have begun to explore ways to make more formal comparisons between the brains of different species. This has been made possible by advances in data acquisition technologies, particularly magnetic resonance imaging (MRI), as well as advances in computing power and data analytic methodology. Thus far, this line of research has primarily explored similarities between the brains of humans and non-human primates. In 2013, Mars et al. first used Passingham's notion of a connectivity fingerprint to identify the macaque homologue of the human temporoparietal junction (Passingham, Stephan, and Kötter 2002; Mars et al. 2013). The connectivity fingerprint is a signature that characterizes how a given region of interest is connected to a set of pre-specified target regions. The targets chosen are areas for which there is strong evidence for homology between the species of interest. Since each region can be uniquely identified by its connectivity fingerprint, it is possible to match fingerprints across species to examine novel patterns of similarity. Mars et al. generated their connectivity fingerprints using resting-state functional MRI (rs-fMRI) data. By building these fingerprints for voxels in the human temporoparietal junction and matching them to the fingerprints of voxels in the macaque brain, they were able to determine which macaque region was most similar. Since then it has been shown that connectivity fingerprints can be used to perform a number of different cross-species comparisons, including matching the fingerprints for a set of pre-specific regions of interest in one species to a template of the brain in the other species, and matching a connectivity fingerprint across a specific region of the brain (e.g. the cortex) in order to identify the closest cross-species match (Mars et al. 2016). In 2018, Mars et al. extended the idea of a connectivity fingerprint to a whole-brain connectivity blueprint, in which a connectivity fingerprint is generated for every region in the brain (Rogier B Mars et al. 2018). In this case, rather than building the connectivity fingerprints using functional connectivity to pre-specified homologous target regions, they used a given region's connection to white matter tracts common among all higher primates. While this line of research has been focused primarily on cross-species comparisons between humans and non-human primates, the development of novel methods has encouraged neuroscientists to think in new ways about exploring cross-species homologies. More recently, Balsters et al. applied the concept of connectivity fingerprints to probe the degree of similarity between the striata of mice, macaques and humans. Their findings suggest that the nucleus accumbens is well conserved across the three species, but that the human striatum contains a large number of voxels with no obvious match in the mouse (Balsters et al. 2020).

Together these results highlight a framework for making formal comparisons between the brains of different species (Rogier B. Mars, Passingham, and Jbabdi 2018; Mars, Jbabdi, and Rushworth 2021). Rather than relying on the definition of homologous neuroanatomical pairs, the degree of similarity between voxels or regions in the brain is quantified directly using an intermediate common space. In order to serve as a bridge between the different species, the common space is constructed using quantitative maps of some underlying homologous biological feature. For instance, in the case of rs-fMRI data (Mars et al. 2013), the space is built using a seed region's functional connections to a set of homologous target regions. With white

matter tractography data (Rogier B Mars et al. 2018), the space is built using a region's connections to homologous white matter tracts. While cross-species comparisons can be accomplished using connectivity profiles, connectivity maps are by no means the only kind of data that can be used to define a common space. In particular, the availability of whole-brain spatial transcriptomic data sets provides an opportunity to build such a space using the expression patterns of homologous genes (Lein et al. 2007; Hawrylycz et al. 2012). This is advantageous for mouse-human comparisons, since the absence of large white matter renders it difficult to make comparisons using structural connectivity. Spatial transcriptomic data contains a wealth of information about neuroanatomical organization, as evidenced by a recent study detailing a novel molecular atlas of the mouse brain (Ortiz et al. 2020). Additionally, past studies have demonstrated that certain broadly defined regions in the mouse and human brains exhibit similarity on the basis of their gene expression profiles (Strand et al. 2007; Myers 2017).

Here we examine the patterns of similarity between the mouse and human brain using a common space constructed from spatial transcriptomic data sets (**Yohan comment to resolve: motivate gene expression**). We begin with an initial set of 2624 homologous genes. Subsequently, we present and evaluate a novel method for improving the resolution of mouse-human neuroanatomical correspondences using supervised machine learning. (**Finish this. Cortex stuff. Striatum stuff. Many to many maps**)

## Results

### Homologous genes capture broad similarities in the mouse and human brains

We first examined the pattern of similarities that emerged when comparing mouse and human brain regions on the basis of their gene expression profiles. We constructed a gene expression common space using widely available data sets from the Allen Institute for Brain Science: the Allen Mouse Brain Atlas (AMBA) and the Allen Human Brain Atlas (AHBA) (Lein et al. 2007; Hawrylycz et al. 2012). These data sets provide whole-brain coverage of expression intensity for thousands of genes in the mouse and human genomes. For our purposes we filtered these gene sets to retain only mouse-human homologous genes using a list of orthologues obtained from the NCBI HomoloGene system (NCBI 2018) (**Tried regenerating this list using Armin's code but the database was outdated and data unavailable**). Prior to analysis, we ran both data sets through a pre-processing pipeline that included quality control checks, normalization procedures, and aggregation of the expression values under a set of atlas labels. The result was a gene-by-region matrix in either species, describing the normalized expression of 2624 homologous genes across 67 mouse regions and 88 human regions (see Materials and methods). We quantified the degree of similarity between all pairs of mouse and human regions using the Pearson correlation coefficient, resulting in a mouse-human similarity matrix (Figure 1, panel A).

We find that the similarity matrix exhibits broad patterns of positive correlation values between the mouse and human brains. These clusters of similarity correspond to coarse neuroanatomical regions that are generally well-defined in both species. For instance, we observe that, overall, the mouse isocortex is similar to the human cerebral cortex, with the exception of the hippocampal formation, which forms a unique cluster. Similarly the mouse and human cerebellar hemispheres cluster together, while the cerebellar nuclei are highly correlated to each other ( $r = 0.404$ ) and also to hindbrain structures like the pons ( $r = 0.359$  and  $r = 0.371$  for the mouse and human nuclei respectively) and myelencephalon ( $r = 0.318$  and  $r = 0.374$ ).

The associations between broad regions such as these are self-evident in the correlation matrix. However the ability to resolve regional matches on a finer scale is limited when using all homologous genes in this way. This is especially true for regions within the cerebral and cerebellar cortices, which exhibit a high degree of internal homogeneity. For example, the variance of the correlations across cortical regions is  $\sigma^2 = 0.0052$  while that across cerebellar hemispheric regions is  $\sigma^2 = 0.0017$ , compared with a total variation of  $\sigma^2 = 0.0416$  across all entries in the matrix. This homogeneity is apparent in the similarity profiles (defined here as the set of correlation values between a given seed region and all target regions in the other species) of the human precentral gyrus and cuneus, both of which exhibit a plateau of similarity to the mouse isocortex (Figure 1, panels B and C). While the brain maps feature a rostral-caudal gradient, the profiles of the two seeds are highly similar despite the regions having vastly different functions. Indeed, the correlation between the similarity profiles of the precentral gyrus and cuneus is  $r = 0.980$ . The similarity profile of human crus 1 highlights another example of this homogeneity. The crus 1 is equivalently similar to all regions of the mouse cerebellum, with an average correlation of  $r = 0.269$  and a standard deviation of  $\sigma = 0.041$ . These findings are in line with what has previously been reported by Myers (Myers 2017) using the same data sets. Examining the correlation between microarray samples in the AHBA and voxels in the AMBA, Myers reported homogeneous patterns of similarity between 11 broadly defined regions of the mouse and human brains. In particular, voxels and samples from the cerebral cortex, the cerebellar hemispheres, and the striatum showed a large degree of internal homogeneity. Importantly, her use of these data at the highest available resolution, i.e. samples and voxels, did not improve the granularity of the neuroanatomical matches beyond what we were able to capture in our regional analysis. This suggests that the regional expression patterns of mouse-human homologous genes can be used to identify general similarities between the brains of the two species even using a simple correlation measure, but the ability to identify finer scale matches might require a more subtle approach.

(Armin suggestion: Include a Sankey diagram for top matches in Figure 1)

(Rogier suggestion: Corresponding mouse and human regions have different colours in Figure 1)

## Classifying mouse neuroanatomy using homologous genes improves the resolution of mouse-human associations

Using the spatial expression patterns of homologous genes, we observed clusters of similarity between broadly defined regions of the mouse and human brains. While the distribution of correlation values within these broad regions is mostly homogeneous, the data show hints of local, i.e. sub-regional, variation, suggesting that the set of homologous genes contains information about mouse-human matches on a finer scale (Figure 1, panel C). Previously it was shown that unsupervised data-driven (e.g. weighted gene co-expression network analysis) or annotation-based (e.g. neuron markers) gene set selection improves the correspondence between established mouse-human regional pairs, but only for broadly defined regions (e.g. cortex-cortex) (Myers 2017). Thus rather than attempting to identify gene subsets to improve the resolution of mouse-human matches, we investigated whether we could accomplish this by constructing a new set of variables from combinations of the homologous genes. Our primary goal here was to transform the initial gene space into a new common space that would improve the locality of the matches. However while we sought a transformation that would allow us to recapitulate known mouse-human neuroanatomical homologues, we

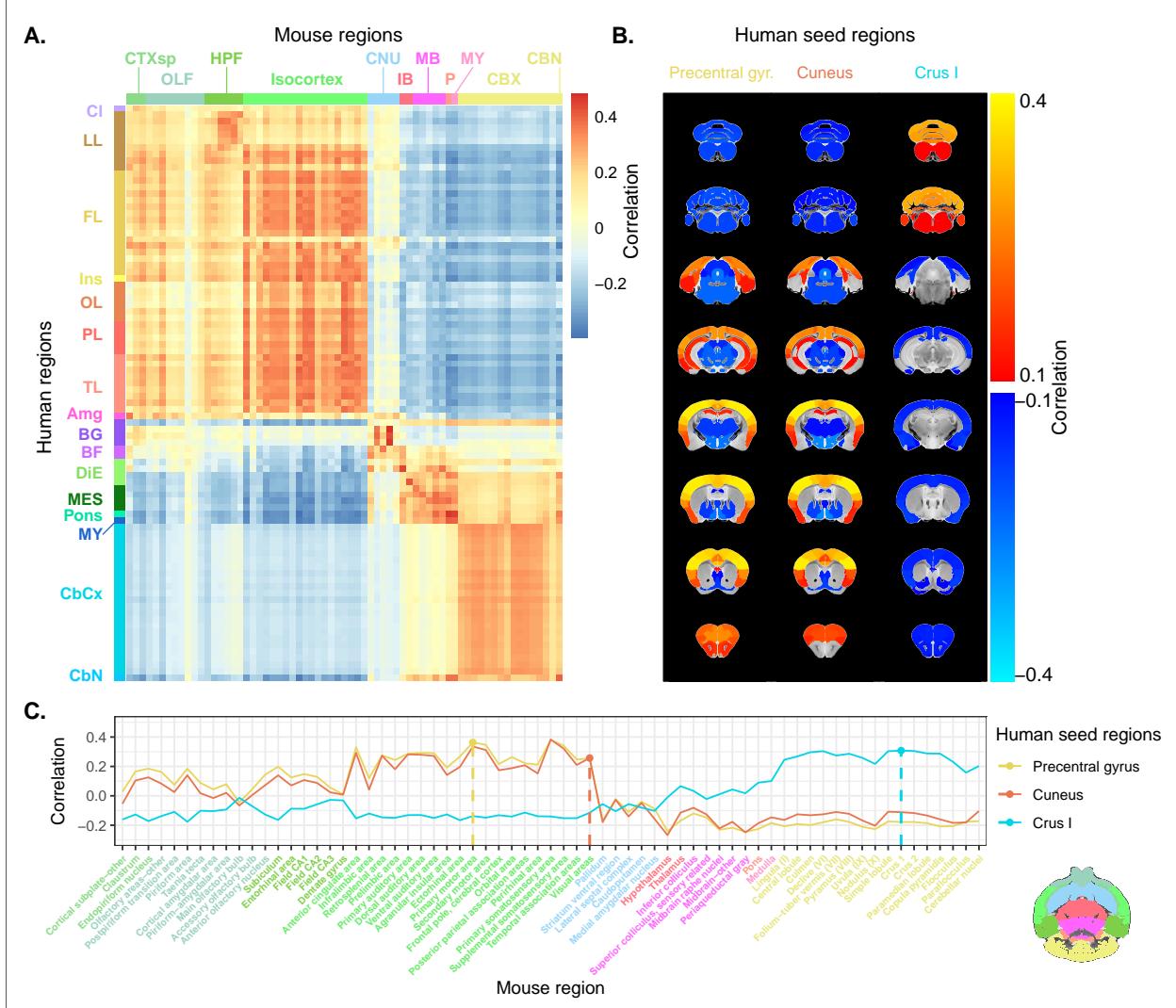


Figure 1: Transcriptomic similarity in the mouse and human brains. (A) Similarity matrix displaying the correlation between 67 mouse regions and 88 human regions based on the expression of 2624 homologous genes. Columns annotated with 11 broad mouse regions: Cortical subplate (CTXsp), olfactory areas (OLF), hippocampal formation (HPF), isocortex, cerebral nuclei (CNU), interbrain (IB), midbrain (MB), pons (P), medulla (MY), cerebellar cortex (CBX), cerebellar nuclei (CBN). Rows annotated with 16 broad human regions: Claustrum (Cl), limbic lobe (LL), frontal lobe (FL), insula (Ins), occipital lobe (OL), parietal lobe (PL), temporal lobe (TL), amygdala (Amg), basal ganglia (BG), basal forebrain (BF), diencephalon (DIE), mesencephalon (MES), pons, myelencephalon (MY), cerebellar cortex (CbCx), cerebellar nuclei (CbN). Broad patterns of similarity are evident between coarsely defined brain regions, while correlation patterns are mostly homogeneous within these regions. (B) Mouse brain slices showing similarity profiles for the human precentral gyrus, cuneus and crus I. Correlation patterns for the precentral gyrus and cuneus are highly similar to one another and broadly similar to most isocortical regions. The crus I is homogeneously similar to the mouse cerebellum. (C) Anatomically-ordered line charts displaying the similarity profiles for the seed regions in (B). Dashed vertical lines indicate the canonical mouse homologue for each human seed. Annotation colours correspond to atlas colours from the AMBA and AHBA for mouse and human regions respectively.

also wanted to avoid directly encoding such correspondences in the transformation. Using this information as part of the optimization process for the transformation would run the risk of driving the transformation towards mouse-human pairs that are already known. While we are interested in being able to recover such matches, we are equally interested in identifying novel and unexpected associations between neuroanatomical regions in the mouse and human brains, e.g. one-to-many correspondences. Given these criteria, our approach to identifying an appropriate transformation was to train a multi-layer perceptron (MLP) classifier on the data from the AMBA. The classifier was tasked with predicting the 67 labels in our mouse atlas from the voxel-wise expression of the homologous genes (Figure 2, panel A). While the model could have been trained using the data from either species, we chose to use the AMBA because it provides continuous coverage of the entire brain and is thus better suited to this purpose. In training the MLP to perform this classification task, we effectively optimize the network architecture to identify a transformation from the input gene space to a space that encodes information about the delineation between mouse brain regions. To extract this transformation, we removed the output layer from the trained neural network. The resulting architecture defines a transformation from the input space to a lower-dimensional gene expression latent space. We then applied this transformation to the mouse and human gene-by-region expression matrices to obtain representations of the data in the latent space (Figure 2, panel B). Finally, we used these gene expression latent space matrices to compute the new similarity matrix (Figure 2, panel C). Since the optimization algorithm used to train the MLP features an inherent degree of stochasticity, we repeated this training and transformation process 500 times to generate a distribution of latent spaces and similarity matrices over training runs.

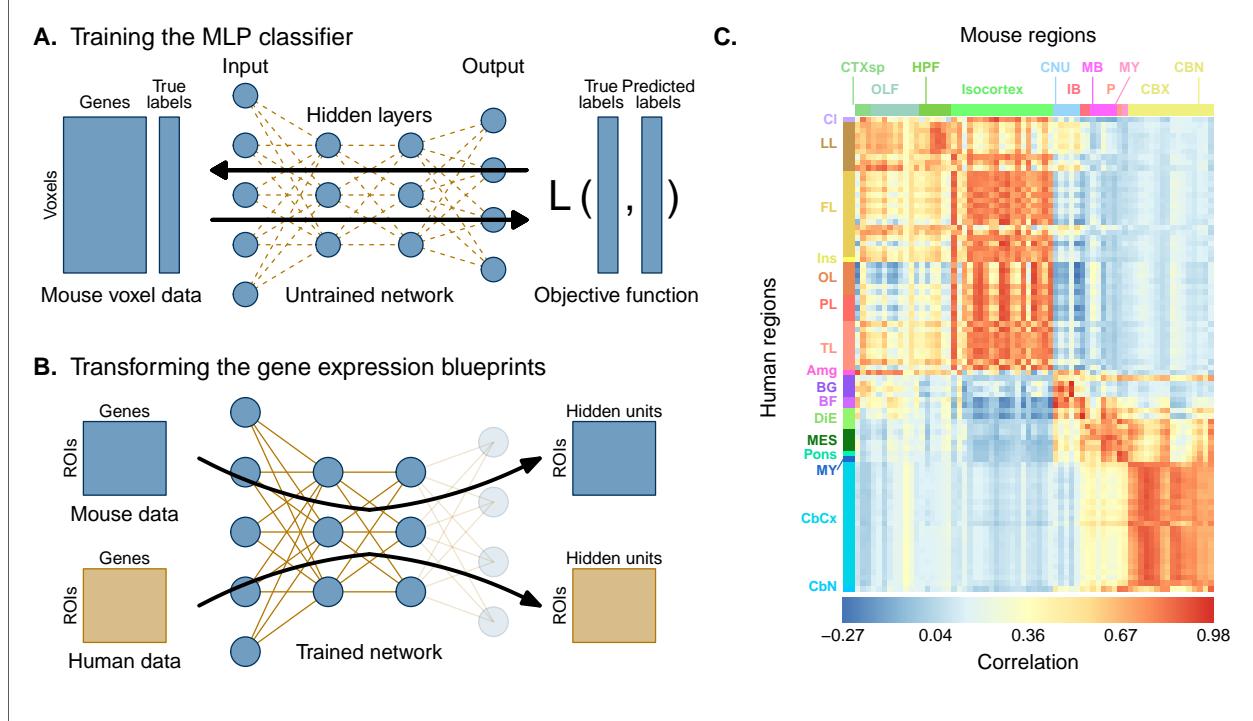


Figure 2: Transforming the common space. (A) Schematic of the neural network training process. Voxel-wise expression maps from 2624 homologous genes in the AMBA were used to train the neural network to classify each mouse voxel into one of 67 atlas regions. (B) Schematic illustrating the transformation from the input space to the gene expression latent space. Once the network is trained, the output layer is removed. The mouse and human regional gene expression matrices are fed through the network, resulting in latent space representations of the data. The training and transformation process was repeated 500 times. (C) Similarity matrix displaying the gene expression latent space correlation between mouse and human regions, averaged over 500 MLP training runs. Similar brain regions exhibit very high correlation values.

To assess whether the latent space representations of the data improved the resolution of the mouse-human matches, we considered two criteria. The first was whether the similarity profiles of the mouse atlas regions were more localized within the corresponding broad regions of interest (e.g. primary motor area within isocortex), compared with their similarity profiles in the original gene space. The second criterion was whether the degree of similarity between canonical neuroanatomical homologues improved in this new common space. The former tells us about our ability to extract finer-scale signal in these profiles, while the latter informs us about our ability to recover expected matches in this finer-scale signal. To evaluate these criteria, we computed ranked similarity profiles for every region in the mouse brain, ordered such that a rank of 1 indicates the most similar human region. In addition, given the difference in absolute value between the input gene space and gene expression latent space correlations, we scaled the similarity profiles to the interval [0, 1] in order to make comparisons between the spaces.

We evaluated the locality criterion by examining the decay rate of the head of the similarity profiles. We reasoned that the plateau of similarity to a broad brain region, as seen in the anatomically-ordered similarity matrices and profiles (Figure 1, panels A and C; Figure 2, panel C), would correspond to a similar plateau at the head of the rank-ordered profiles. Moreover, the emergence of local signal would manifest as an increase in the range between the peaks and troughs within the broad region. In the rank-ordered profiles, this would correspond to a faster rate of decay at the head of the profile. In order to quantify this decay, we computed the rank at which each region's similarity profile decreased to a scaled value of 0.75 (**Yohan comment: How was this chosen? Are the results stable with respect to this parameter?**). This was calculated for every mouse region in the initial gene space, as well as in each of the 500 gene expression latent spaces. As a measurement of performance between the two representations of the data, we then took the difference in this rank between each of the latent spaces and the original gene space (Figure 3, panel A). A negative rank difference indicates an improvement in the latent space.

#### (Armin comment 1: Possible to extract gene weights per ROI?)

Examining the structure-wise distributions of these rank differences, we found that for the majority of regions in our mouse atlas, the classification approach resulted in either an improvement in the amount of locality within a broad region, or no difference from the original gene space (Figure 3, panels B and C). Specifically, 47 regions (70.1%) had a mean rank difference less than or equal to zero. Additionally, the same number of regions returned non-positive rank differences in at least 80% of latent spaces. A few regions performed considerably worse in the latent spaces, notably the main olfactory bulb ( $\mu = 18.4$ ;  $\sigma = 12.7$ ), the accessory olfactory bulb ( $\mu = 8.7$ ;  $\sigma = 11.6$ ), and the cerebellar nuclei ( $\mu = 9.1$ ;  $\sigma = 8.5$ ) (**Armin comment 2, Yohan comment 4: Why?**). In particular, the main olfactory bulb performed worse in 96.6% of latent spaces. Regions within the cortical subplate and olfactory areas (e.g. endopiriform nucleus, postpiriform transition area) benefited the most from the classification approach, with many regions showing improvements in all latent spaces. While the effects are smaller, the similarity profiles of regions belonging to the isocortex and cerebellar cortex also saw an improvement in locality. In the isocortex, 16 out of 19 regions (84.2%) improved in at least 96% of latent spaces. In the cerebellar cortex, 73.3% of regions saw a similar improvement. In contrast, regions belonging to the cerebral nuclei, the diencephalon, midbrain and hindbrain did not see much improvement in this new common space. For instance, only 13.2% of latent spaces returned a non-positive rank difference in the thalamus. For many such regions the degree of locality appears to be worse in this space, though only by a small number of ranks (e.g. striatum ventral region, thalamus, midbrain raphe nuclei). Indeed, the mean rank difference and standard deviation over these regions and all latent spaces

are  $\mu = 1.4$  and  $\sigma = 3.6$ . These results demonstrate that the supervised learning approach used here can improve the resolution of neuroanatomical correspondences between the mouse and human brains, though the amount of improvement varies over the brain. Regions that were already well-characterized using the initial set of homologous genes (e.g. subcortical regions) did not benefit tremendously, but numerous regions in the cortical plate and subplate, as well as the cerebellum, saw an improvement in locality in this new common space.

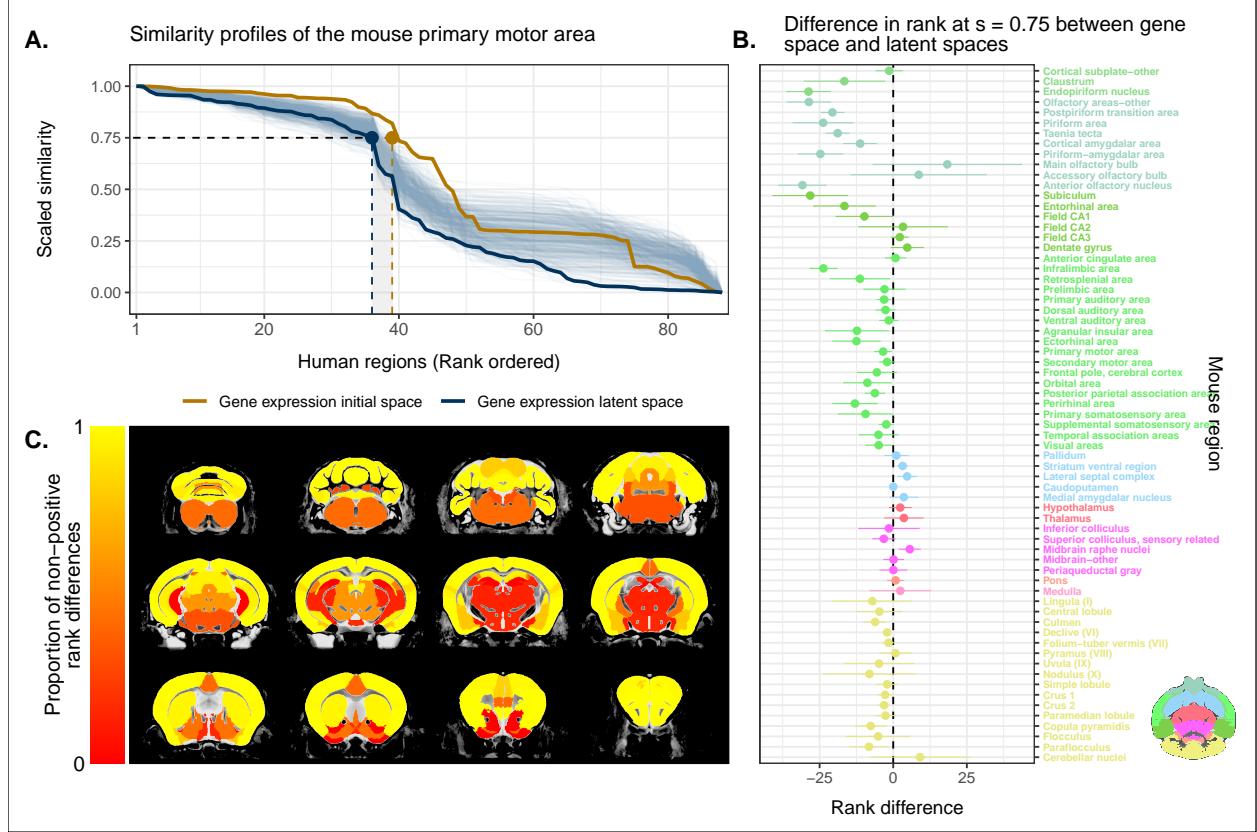


Figure 3: Quantifying improvement in locality in gene expression latent space. (A) The amount of local signal within a broadly similar region of the brain for a finer seed region's (e.g. primary motor area) similarity profile can be quantified by the decay rate of the head of the rank-ordered profile. Decay rate was quantified by computing the rank at a similarity of 0.75. This metric was compared between the initial gene expression space (orange line) and every gene expression latent space resulting from repeated training of the MLP (every blue line is a training outcome, heavy blue line serves as an example). A negative difference between these rank metrics indicates an improvement in locality in the latent space. (B) Structure-wise distributions of differences in rank at a similarity of 0.75 between the initial gene expression space and the gene expression latent spaces. Points and error bars represent mean and 95% confidence interval. Dashed black line at 0 indicates the threshold for improvement in one space over the other. Colours correspond to AMBA annotations as in Figures 1 and 2. (C) Proportion of MLP training runs resulting in an improvement or null difference in the gene expression latent space compared with the initial space. Cortical and cerebellar regions exhibit high proportions of improvement, while subcortical regions are less likely to be improved by the classification process.

While the supervised learning approach improved our ability to identify matches on a finer scale for a number of brain regions, this does not necessarily mean that those improved matches are biologically meaningful. Our second criterion for evaluating the performance of the neural network addresses whether this improvement in locality captures what we would expect in terms of known mouse-human homologies. To this end, we examined the degree of similarity between established mouse-human neuroanatomical matches, both in the initial gene expression space and in the set of latent spaces. We began by establishing a list of 37 canonical mouse-human homologous pairs. For each of these regions in the mouse brain, we evaluated how the rank of the canonical human match changed in the rank-ordered similarity profiles between the latent spaces and

the original gene space (Figure 4, panel A). The lower the rank, the more similar the canonical pair, with a rank of 1 indicating maximal similarity. We additionally calculated the proportion of latent spaces in which each mouse region was more similar or as similar to its canonical human match compared with the initial gene space (Figure 4, panel B). We find that for most regions in the mouse brain, the classification approach either improves the correspondence or performs as well as the full set of homologous genes. For example, 73% of regions exhibit improved similarity in at least 80% of latent spaces. The improvement is most pronounced for regions in the cortical subplate and isocortex. In particular, the frontal pole improves from a rank of 33 to an average rank of 3. Similarly, the visual areas improve from a rank of 32 to an average of 10, though the variance is much higher in this case. Many regions in the sub-cortex don't benefit from the gene expression latent spaces since the initial gene set was already recapitulating the appropriate match with maximal similarity. Apart from the pallidum and the medulla, each of these regions is maximally similar to its canonical match in at least 90% of latent spaces. In such cases, the classification approach performs as well as the original approach. Finally, while many regions in the cerebellum feature some improvement in the latent spaces, the variation in the rank of the standard human pair is often quite large, indicating some instability in the neural network's ability to recover these matches. To understand this pattern, we examined the latent space distributions of the top 24 matches for cerebellar cortical regions such as the crus 1 and copula pyramidis (Figure 4, panel C). We find that these matches are always cerebellar regions. Thus the high degree of variance observed in the ranks of the canonical cerebellar pairs corresponds to a shuffling of the cerebellar targets on the human side. For instance, when the mouse crus 1 is used as the seed region, the human crus 1 is most often assigned a rank between 6 and 9. However, similar proportions in that range occur for the crus 2 and lobules V, VI and VIIIB, indicating that these cerebellar regions are swapping ranks in the different latent spaces. Thus while cerebellar regions are reliably associated with other cerebellar regions in the gene expression latent spaces, these associations are not stable over multiple training runs.

(**Elaborate on how cerebellar regions are hard to distinguish using gene expression**)

(**Finish organizing Figure 4**)

These results demonstrate that the MLP classification approach improves our ability to resolve finer scale mouse-human neuroanatomical matches within the broadly similar regions obtained using the initial gene expression space. By training a classifier to predict the atlas labels in one species, we were able to generate a new common space that amplified the amount of local signal within broadly similar regions while also improving our ability to recover known neuroanatomical homologues.

## Sensorimotor areas of the mouse isocortex are more strongly conserved than association areas

Having validated our ability to resolve finer scale matches using the gene expression latent space, we proceeded to investigate hypotheses about conservation of brain areas between the mouse and human. We hypothesized that sensorimotor areas of the cortex would be more conserved than association areas (**Need some literature justification and contextualization here**). We quantified the degree of conservation for a given cortical seed region in one species as the maximal correlation value across all target regions in the other species. We evaluated this quantity for every region in the mouse isocortex in each of the 500 gene expression latent spaces (Figure 5, panel A). We further probed the association between cortex type and conservation by computing the average maximal correlation per cortex type in each of the gene expression la-

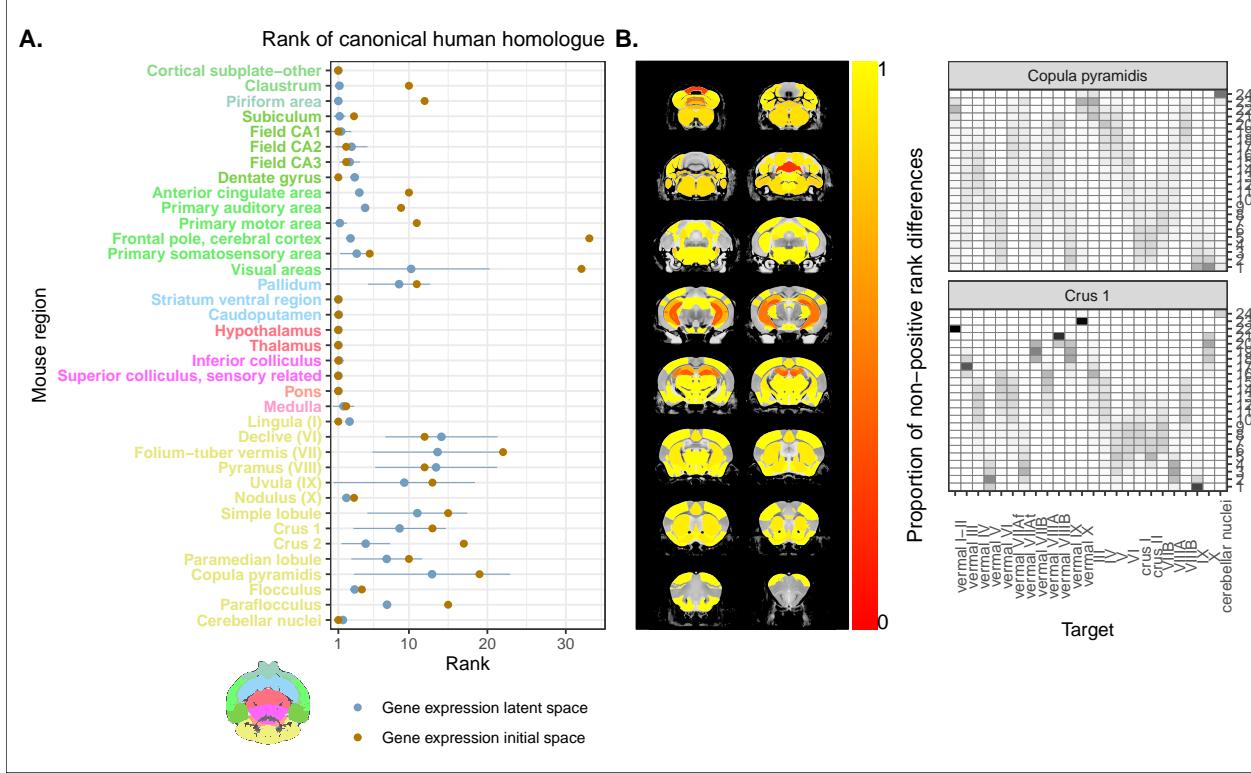


Figure 4: Recovering canonical neuroanatomical pairs in gene expression space. (A) Comparison between the ranks of canonical human matches for mouse seed regions between the initial gene expression space and gene expression latent spaces. Points and error bars represent mean and 95% confidence interval. Axis colours correspond to AMBA annotations. (B) Proportion of latent spaces resulting in an improvement or null difference compared with the initial gene space. Uncoloured areas correspond to regions with no established canonical human match. (C)

tent spaces (Figure 5, panel B). On average, we find that mouse sensorimotor cortical areas exhibit a greater degree of conservation than association areas, with the primary motor and somatosensory areas being the most conserved. Despite the stochasticity that emerges due to the training process of the MLP classifier, mouse sensorimotor cortical areas consistently exhibit larger maximal correlation values than association areas.

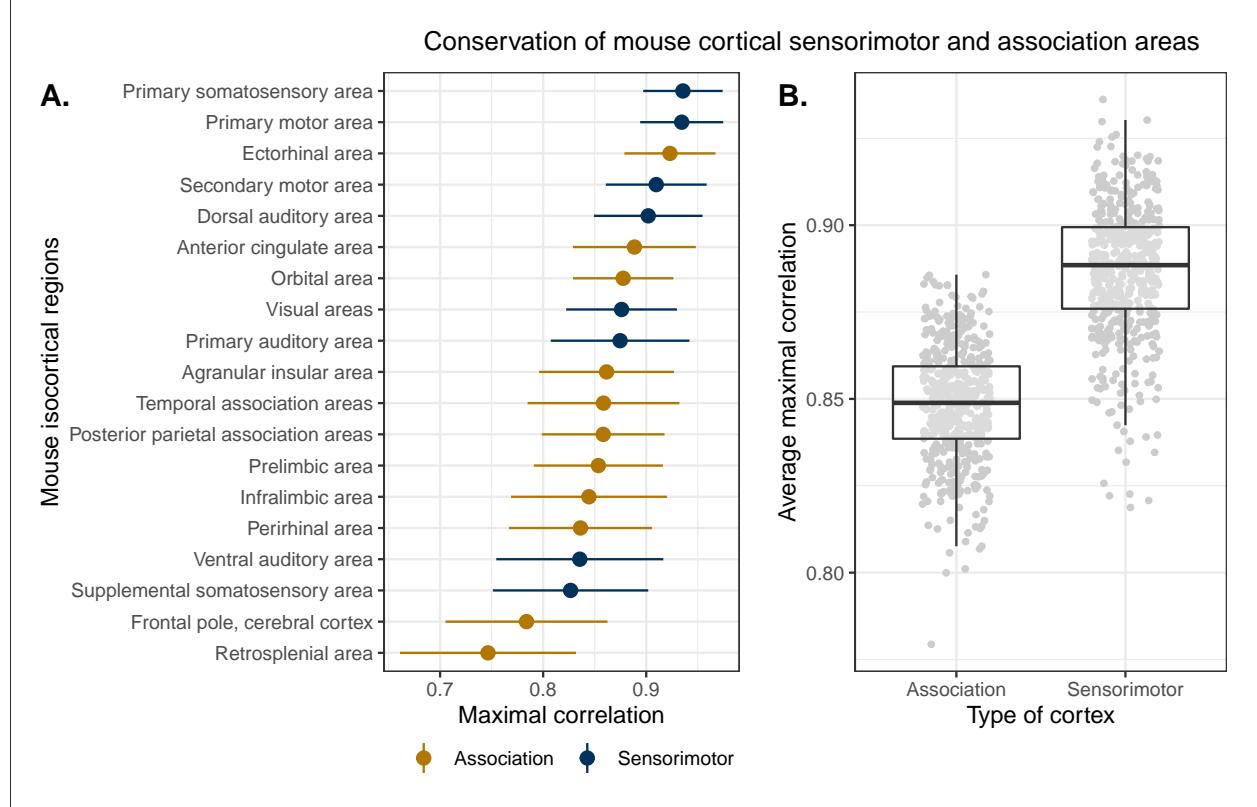


Figure 5: Conservation of mouse cortical areas. The gene expression latent space can be used to evaluate evolutionary hypotheses such as whether sensorimotor or association cortical areas are more conserved between the mouse and human. (A) Conservation was evaluated using the maximal correlation value of seed region. Points and error bars represent mean and 95% confidence interval over latent space samples. Sensorimotor areas tend to have higher maximal correlation values on average. (B) Average maximal similarity per cortex type in each gene expression latent space.

## Results 4

An advantage of the common space approach is that it allows for the quantitative comparison of brain organization between related species. This gives rise to the possibility of exploring more nuanced relationships between neuroanatomical regions. Moreover, common spaces built using diverse features of the brain provide information about multiple facets of brain organization. Given that the majority of studies using the common space approach have focused on comparing primate brains, Balsters et al. (Balsters et al. 2020) were among the first to apply it to mice and humans. Using rs-fMRI, they evaluated the similarity of voxels in the striatum across species. They found that the nucleus accumbens was highly conserved across mice and humans, and that voxels in the posterior part of the human putamen were significantly similar to the lateral portion of their mouse caudoputamen parcellation. Additionally, they report that 85% of voxels in the human striatum were not significantly similar to any of their mouse striatal seeds, and that 25% of human striatal voxels were

significantly *dissimilar* compared with the mouse. Motivated by these results, we investigated the patterns of similarity between the mouse and human striatum (**striata?**) on the basis of gene expression, specifically using the MLP latent space representations.

We first identified the striatal regions present in the AHBA; these were the caudate nucleus, the putamen, and the nucleus accumbens. We evaluated the correlation between the microarray samples in these regions and every region in the mouse atlas. Based on these correlation values, we focused our analysis on the four mouse regions that were consistently the most similar across all latent spaces: the caudoputamen, the nucleus accumbens, the fundus of striatum, and the olfactory tubercle. For each of the human striatal regions, we then calculated the average correlation over the samples to each of the mouse targets. We examined the distribution of these average correlation values over the latent spaces (Figure 6, panel A). We find that the human caudate and putamen consistently exhibit the strongest degree of similarity to the mouse caudoputamen. Beyond this expected top match, the caudate and putamen both exhibit high similarity to the nucleus accumbens and the fundus of striatum. Neither of these target regions is consistently more similar to the mouse caudoputamen over all latent spaces. While the similarity of the caudate and the putamen to the caudoputamen is unsurprising, the story isn't as clear for the human nucleus accumbens. We find that the correlation values are slightly lower overall, indicating less conservation of the nucleus accumbens. We also find that the human nucleus accumbens isn't specifically similar to the mouse nucleus accumbens in the way that the caudate and putamen are similar to the caudoputamen. The similarity of the human accumbens is distributed mostly between the mouse caudoputamen, accumbens, and fundus of the striatum. The human accumbens also exhibits a high degree of similarity to the mouse olfactory tubercle.

The sparse sampling of the microarray survey in the AHBA makes it difficult to examine patterns of similarity at a sub-regional level in the striatum. As such, we chose to examine how voxels in the mouse striatum correspond to human regions. For every voxel in the mouse striatum, we computed the correlation to human regions in our atlas. We then identified the maximally similar human region in each of the latent spaces. We created maps illustrating the proportion of latent spaces in which each voxel is maximally similar to various human regions (Figure 6, panel B). We find that voxels in the caudoputamen are most often maximally similar to the human caudate and putamen, and that the pattern of similarity segments the caudoputamen into two sub-regions based on the target. This pattern exhibits a strong degree of bilaterality. (**Elaborate on this**). Voxels that are most similar to the the human nucleus accumbens are present in both the mouse nucleus accumbens and the olfactory tubercle. Finally, we find that the set of voxels most similar to the septal nuclei belong to the lateral septal complex.

## Discussion

We have demonstrated how the spatial transcriptomic patterns of homologous genes can be used to make quantitative comparisons between the mouse and human brain. We presented a novel approach using supervised learning in which a lower dimensional representation of the data was used to extract mouse-human neuroanatomical correspondences on a finer scale. We found that this approach both enhanced the locality of matches within broadly similar regions, and improved our ability to recapitulate expected mouse-human homologues. Using this method, we confirmed that sensorimotor areas of the mouse isocortex exhibit a greater degree of conservation compared with association areas. We also demonstrated that regions in the mouse and human striata exhibit patterns of correspondence that go beyond one-to-one matches.

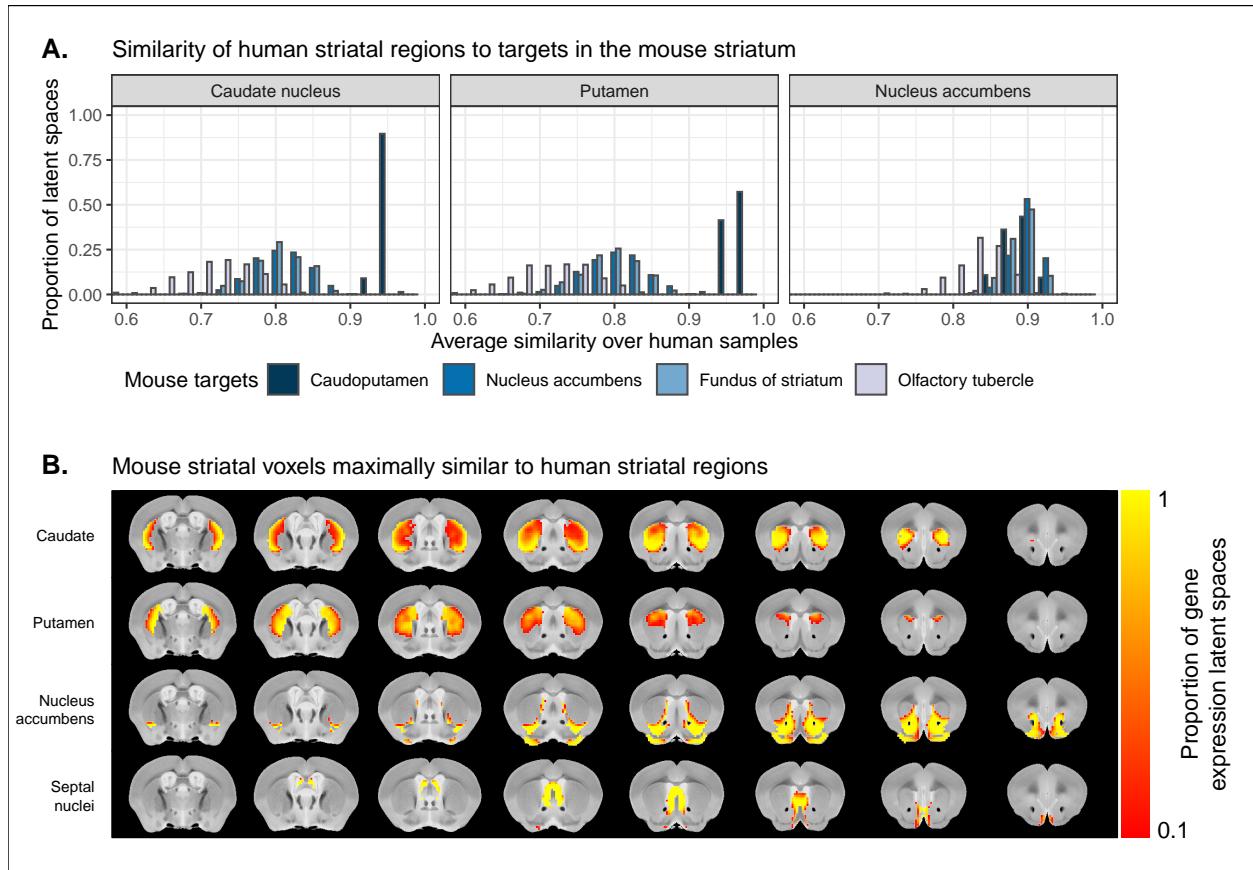


Figure 6: Patterns of similarity between mouse and human striatal regions. (A) Distributions over gene expression latent spaces of region-wise average correlation values for mouse and human striatal pairs. (B) Proportions of latent spaces in which mouse striatal voxels are maximally similar to human target regions.

The abundance of neuroscience research performed using mice has resulted in a wealth of knowledge about the mouse brain. In the preclinical setting, mouse models are utilized with the intention of better understanding human neuropathology. For instance, in the context of autism spectrum disorders, a plethora of studies using mouse models have reported on the neurobiological and neuroanatomical phenotypes that arise from mutations at specific genetic loci. It is common for researchers involved in translational neuroscience to rely on findings of this kind to make inferences about the human disorder and establish some level of face validity for the model. The typical approach is to identify rough correspondences between neuroanatomical ontologies in both species. While this is a reasonable starting point for comparison, the true correspondence between the mouse and human brain is likely more complicated given the evolutionary distance between the two species. We expect that coarsely defined neuroanatomical regions will be conserved to a large extent (**why?**), but many finer scale regions are likely to have diverged and specialized over time as the two species adapted to their specific ecological niches. A consequence of this divergence may be the re-organization of brain regions. It is unlikely that every brain region has a singular correspondence between the two species, suggesting that a one-to-one approach does not capture the entire story. In our analysis of the striatum, we find patterns that exhibit many-to-many correspondences between the mouse and human brain. The mouse caudoputamen, which is typically delineated as a single region, exhibits sub-structural patterns of organization based on the degree of similarity to the human caudate and putamen. Balsters et al. [Balsters et al. (2020) reported a similar finding. We also find a multiplicity of matches when examining the nucleus accumbens, with the human accumbens exhibiting diffuse similarity to a number of mouse striatal regions.

The expression of homologous genes provides an elegant way to define a common space for cross-species comparisons since it relies on homology at a deep molecular biological level. The approach is not without limitations however. The acquisition of spatial transcriptomic data is labour-intensive, time consuming, and invasive. These data sets cannot be generated easily, especially in the human, in which the process depends on the availability of post-mortem samples. As a result, the effective sample sizes are extremely limited in this domain. For instance, in the AMBA coronal data set used here, the brain-wide expression of each gene is sampled only once (barring a few exceptions). This constrains the type of analyses that are possible (e.g. null-hypothesis significance testing) and largely limits the availability of replication data sets. Other modalities such as rs-fMRI are advantageous in this regard. Larger samples can be generated with relative ease in a variety of species. The caveat is that the resulting common space relies on a definition of homology at a higher level, i.e. a priori definitions of neuroanatomical homologues. That being said, both of these modalities are useful in their own right and likely provide complementary information about mouse-human brain correspondences. Used individually, they highlight the different ways in which brain regions correspond to one another. However, moving forward, the availability of similar data sets in mice and humans over a range of modalities opens the door to the possibility of multi-modal cross-species comparisons. One possible approach is the application of data integration strategies such as similarity network fusion to generate a multi-modal similarity network between mouse and human brain regions or voxels.

The common space framework is a powerful way to explore and evaluate similarity between the brains of different species. This is especially true for species as distantly related as the mouse and human, where methods such as cross-species image registration or cortical expansion maps are not feasible. While the use of this framework for mouse-human brain comparisons is still in its infancy, it shows promise for elucidating the relationship between mouse and human brains and improving our understanding of the translational potential between these two species. Importantly, it also provides an avenue for new ways of translating results between the species, such as the direct transformation of brain-wide maps from one species to the

other.

## Materials and methods

### Mouse gene expression data

For the mouse expression data, we used the whole brain in-situ hybridization (ISH) coronal data set from the Allen Mouse Brain Atlas (Lein et al. 2007). (**Sentence or two about how ISH MINC files were generated from the raw AMBA data**). The ISH images for 4346 genes were imported and masked to form a gene-by-voxel expression matrix. We pre-processed this data by first applying a  $\log_2$  transformation for consistency with the human data set. Subsequently we filtered out genes for which more than 20% of voxels contained missing values. For those genes with replicated ISH experiments, we averaged expression of each voxel across the experiments. Finally we applied a K-nearest neighbours algorithm to impute the remaining missing values. This was done using the genes as the variables. The result was a gene-by-voxel expression matrix with 3958 genes and 61316 voxels.

### Human gene expression data

Human gene expression data was obtained from the Allen Human Brain Atlas (Hawrylycz et al. 2012). We used the microarray data from the brains of all 6 donors, each of which contains  $\log_2$  expression values for 58692 gene probes across numerous tissue samples. The data sets were fed through a pre-processing pipeline built following the recommendations from Arnatkevičiūtė et al. (2019) (Arnatkevičiūtė, Fulcher, and Fornito 2019). Specifically, once imported, we passed the data from each donor through a set of filters. The first filter removed gene probes that were not associated with an existing Entrez gene ID. The second filtering step used the probe intensity filter provided by the AHBA. For each donor, we only retained the probes for which more than 50% of samples passed the intensity filter. After filtering, we aggregated the expression values for probes that corresponded to the same gene. To do so, we computed the average expression per sample for probes corresponding to a given gene. This was done separately for each donor, and the averages were computed in linear space rather than  $\log_2$  space. Once the average gene expression values were obtained, we transformed the data back to  $\log_2$  space. Finally, we combined the gene-by-sample expression matrices across the different donors. In doing so, we retained only those genes present in the data sets from all 6 donors. The result was a gene-by-sample expression matrix with 15125 genes and 3703 samples.

### Mouse atlases

We used a modified version of the DSURQE atlas (Dorr et al. 2008; Richards et al. 2011; Ullmann et al. 2013; Steadman et al. 2014; Qiu et al. 2018) at the Mouse Imaging Centre. The labels of the atlas are mapped to the finest regions in the AMBA ontology, resulting in a hierarchical neuranatomical tree. We used this tree to prune the hierarchy to the desired level of granularity. White matter and ventricular segmentations were removed entirely. The remaining grey matter regions were pruned so that the majority of leaf nodes contained enough voxels to be classified appropriately by the multi-layer perceptron. In doing so, we also attempted to maintain the same level of tree depth within a broad region, e.g. cerebellar regions

were chosen at the same level of granularity. The leaf nodes of this tree were used as the labels in our custom atlas, resulting in a mouse atlas with 67 grey matter regions. We additionally generated an atlas with 11 broader regions for visualization and annotation purposes.

## Human atlases

For our human atlas we used the ontology from the AHBA. We pruned the neuroanatomical hierarchy to correspond roughly to the level of granularity in our mouse atlas, resulting in 88 human brain regions. We additionally generated a set of 16 broad regions for visualization and annotation. White matter and ventricular regions were omitted entirely.

## Creating similarity matrices

We created the mouse and human gene expression blueprints from the human gene-by-sample expression matrix and the mouse gene-by-voxel expression matrix. We first intersected the gene sets in these matrices using a list of homologous genes obtained from the NCBI HomoloGene database. The resulting set contained 2624 homologous genes for both the mouse and human data. We then annotated each of the human samples with one of the 88 human regions in our atlas. The same was done for the mouse voxels with the 67 mouse atlas regions. Voxels and samples corresponding to white matter and ventricular regions were discarded in this process. These labelled expression matrices were subsequently normalized. For each matrix, we first standardized every gene across all voxels or samples. We then centered every voxel or sample across all genes. Finally, we generated the expression blueprints by averaging the expression of every homologous gene over the voxels or samples that corresponded to each atlas region. Using these blueprints, we computed the mouse-human similarity matrix by computing the Pearson correlation coefficient between all pairs of mouse and human regions.

## Validating the multi-layer perceptron

Prior to training the multi-layer perceptron classifier, we applied a validation procedure to tune the hyperparameters of the network. We chose a model architecture in which each layer of the network was fully connected to the previous and subsequent layers. However we still had flexibility in setting a number of hyperparameters, namely the number of hidden layers in the network, the number of hidden units (i.e. neurons) per hidden layer, the dropout rate, and the amount of weight decay. Given that the majority of genes in the coronal AMBA data set are only sampled once over the entire mouse brain, we devised an ad hoc validation procedure to estimate the out-of-sample error of our model as we varied these hyperparameters. The process involved a combination of the coronal AMBA data set and the sagittal AMBA data set. Since the sagittal data set features more genes than the coronal data set, we first filtered the list of genes for those present in the coronal set. We then imported and pre-processed the sagittal data using the process described above for the coronal data set. Since the sagittal data set only covers one hemisphere of the mouse brain, we also repeated the importation and pre-processing process for the coronal data, but masking it using the sagittal brain mask so that only one hemisphere was retained. In applying this pipeline, we did not aggregate the expression of multiple ISH experiments for those genes featuring more than one experiment. Thus some

of the genes in this coronal set had more than one sample. Once these gene-by-voxel expression matrices were built, we filtered each of them according to the list of mouse-human homologous genes and the human sample expression matrix. We then annotated the voxels in each of the mouse voxel expression matrices with one of the 67 regions in our atlas. In order to validate the MLP, we had to generate both a training set and a validation set. To do so, we applied the following procedure: For every gene in the homologous set, we first determined whether the gene had more than one sample in the coronal ISH data set. If this was the case, we randomly chose one of those samples for the training set and one of the remaining samples for the validation set. However if the gene only had a single sample in the coronal data set, we randomly chose either the coronal sample or the sagittal sample for the training set and the other for the validation set. Once the training and validation sets were generated, we trained the MLP on the training set and evaluated its performance on the validation set. Both the training and validation sets were normalized as described above prior to training. We repeated this entire procedure 10 times for every hyperparameter combination. We varied the number of hidden layers between 3 and 5. We varied the number of hidden units between 100 and 1000, increasing them in increments of 100. For the dropout rate, we examined values of 0, 0.25 and 0.50. For the amount of weight decay, we examined values of  $10^{-6}$  and  $10^{-5}$ . Following this validation procedure, we found that the best performing model had 3 hidden layers, 200 neurons per layer, a dropout rate of 0, and a weight decay value of  $10^{-6}$ . This model had an average accuracy of 0.926 on the training sets and of 0.526 on the validation sets.

## Training the multi-layer perceptron

Following validation, we trained the MLP classifier on the full coronal ISH data set from the AMBA. The voxel-wise expression matrices were labelled and normalized as described previously. The hyperparameters of the network were set to the optimal values obtained from the validation procedure. The network was trained using a learning rate of  $10^{-5}$  over 200 epochs. (**Confirm what the loss function was for this and the validation procedure. Maybe also mention activation function, optimizer, etc.**).

## Creating the gene expression latent space

The trained MLP was used to generate the new gene expression latent space. To do so, we removed the predictive output layer from the network architecture, thus allowing the hidden units of the third hidden layer to be the output of the network. We then applied this modified network to the mouse and human gene expression blueprints, resulting in matrices describing each brain region's weight over 200 hidden units. These matrices form the blueprints in the latent space.

## References

Arnatkevičiūtė, Aurina, Ben D. Fulcher, and Alex Fornito. 2019. “A Practical Guide to Linking Brain-Wide Gene Expression and Neuroimaging Data.” *NeuroImage* 189 (April): 353–67. <https://doi.org/10.1016/j.neuroimage.2019.01.011>.

Balsters, Joshua Henk, Valerio Zerbi, Jerome Sallet, Nicole Wenderoth, and Rogier B Mars. 2020. “Primate Homologs of Mouse Cortico-Striatal Circuits.” *eLife* 9 (e53680): 24. <https://doi.org/10.7554/eLife.53680>.

- Bronchti, Gilles, Peter Heil, Ronen Sadka, Andreas Hess, Henning Scheich, and Zvi Wollberg. 2002. "Auditory Activation of 'Visual' Cortical Areas in the Blind Mole Rat (*Spalax Ehrenbergi*)." *European Journal of Neuroscience* 16 (2): 311–29. <https://doi.org/10.1046/j.1460-9568.2002.02063.x>.
- Dietrich, Michael R, Rachel A Ankeny, and Patrick M Chen. 2014. "Publication Trends in Model Organism Research." *Genetics* 198 (3): 787–94. <https://doi.org/10.1534/genetics.114.169714>.
- Dorr, A.E., J.P. Lerch, S. Spring, N. Kabani, and R.M. Henkelman. 2008. "High Resolution Three-Dimensional Brain Atlas Using an Average Magnetic Resonance Image of 40 Adult C57Bl/6J Mice." *NeuroImage* 42 (1): 60–69. <https://doi.org/10.1016/j.neuroimage.2008.03.037>.
- Ellenbroek, Bart, and Jiun Youn. 2016. "Rodent Models in Neuroscience Research: Is It a Rat Race?" *Disease Models & Mechanisms* 9 (10): 1079–87. <https://doi.org/10.1242/dmm.026120>.
- Hawrylycz, Michael J., Ed S. Lein, Angela L. Guillozet-Bongaarts, Elaine H. Shen, Lydia Ng, Jeremy A. Miller, Louie N. van de Lagemaat, et al. 2012. "An Anatomically Comprehensive Atlas of the Adult Human Brain Transcriptome." *Nature* 489 (7416): 391–99. <https://doi.org/10.1038/nature11405>.
- Hedrich, Hans J, Horst Mossmann, and Werner Nicklas. 2004. "Chapter 24: Housing and Maintenance." In *The Laboratory Mouse*, 1st Edition, 395–408. Elsevier Academic Press.
- Houdebine, Louis-Marie. 2004. "Chapter 6: The Mouse as an Animal Model for Human Diseases." In *The Laboratory Mouse*, 1st Edition, 97–107. Elsevier Academic Press.
- Kaas, J. H. 1983. "What, If Anything, Is SI? Organization of First Somatosensory Area of Cortex." *Physiological Reviews* 63 (1): 206–31. <https://doi.org/10.1152/physrev.1983.63.1.206>.
- Kaas, Jon H. 2005. "From Mice to Men: The Evolution of the Large, Complex Human Brain." *Journal of Biosciences* 30 (2): 155–65. <https://doi.org/10.1007/BF02703695>.
- Krubitzer, Leah, and Jon Kaas. 2005. "The Evolution of the Neocortex in Mammals: How Is Phenotypic Diversity Generated?" *Current Opinion in Neurobiology* 15 (4): 444–53. <https://doi.org/10.1016/j.conb.2005.07.003>.
- Lein, Ed S., Michael J. Hawrylycz, Nancy Ao, Mikael Ayres, Amy Bensinger, Amy Bernard, Andrew F. Boe, et al. 2007. "Genome-Wide Atlas of Gene Expression in the Adult Mouse Brain." *Nature* 445 (7124): 168–76. <https://doi.org/10.1038/nature05453>.
- Mars, R. B., J. Sallet, F.-X. Neubert, and M. F. S. Rushworth. 2013. "Connectivity Profiles Reveal the Relationship Between Brain Areas for Social Cognition in Human and Monkey Temporoparietal Cortex." *Proceedings of the National Academy of Sciences* 110 (26): 10806–11. <https://doi.org/10.1073/pnas.1302956110>.
- Mars, Rogier B., Saad Jbabdi, and Matthew F.S. Rushworth. 2021. "A Common Space Approach to Comparative Neuroscience." *Annual Review of Neuroscience* 44 (1): 69–86. <https://doi.org/10.1146/annurev-neuro-100220-025942>.
- Mars, Rogier B., Richard E. Passingham, and Saad Jbabdi. 2018. "Connectivity Fingerprints: From Areal Descriptions to Abstract Spaces." *Trends in Cognitive Sciences* 22 (11): 1026–37. <https://doi.org/10.1016/j.tics.2018.08.009>.
- Mars, Rogier B, Stamatios N Sotiroopoulos, Richard E Passingham, Jerome Sallet, Lennart Verhagen, Alexandre A Khrapitchev, Nicola Sibson, and Saad Jbabdi. 2018. "Whole Brain Comparative Anatomy Using Connectivity Blueprints." *eLife* 7 (e35237). <https://doi.org/10.7554/eLife.35237>.

- Mars, Rogier B., Lennart Verhagen, Thomas E. Gladwin, Franz-Xaver Neubert, Jerome Sallet, and Matthew F.S. Rushworth. 2016. "Comparing Brains by Matching Connectivity Profiles." *Neuroscience & Biobehavioral Reviews* 60 (January): 90–97. <https://doi.org/10.1016/j.neubiorev.2015.10.008>.
- Myers, Emma. 2017. "Molecular Neuroanatomy: Mouse-Human Homologies and the Landscape of Genes Implicated in Language Disorders." PhD thesis, Boston University.
- NCBI. 2018. "Database Resources of the National Center for Biotechnology Information." *Nucleic Acids Research* 46 (Database issue): D8–D13. <https://doi.org/10.1093/nar/gkx1095>.
- Ortiz, Cantin, Jose Fernandez Navarro, Aleksandra Jurek, Antje Märtin, Joakim Lundeberg, and Konstantinos Meletis. 2020. "Molecular Atlas of the Adult Mouse Brain." *Science Advances* 6 (eabb3446): 14.
- Passingham, Richard E., Klaas E. Stephan, and Rolf Kötter. 2002. "The Anatomical Basis of Functional Localization in the Cortex." *Nature Reviews Neuroscience* 3 (8): 606–16. <https://doi.org/10.1038/nrn893>.
- Qiu, Lily R., Darren J. Fernandes, Kamila U. Szulc-Lerch, Jun Dazai, Brian J. Nieman, Daniel H. Turnbull, Jane A. Foster, Mark R. Palmert, and Jason P. Lerch. 2018. "Mouse MRI Shows Brain Areas Relatively Larger in Males Emerge Before Those Larger in Females." *Nature Communications* 9 (1): 2615. <https://doi.org/10.1038/s41467-018-04921-2>.
- Richards, Kay, Charles Watson, Rachel F. Buckley, Nyoman D. Kurniawan, Zhengyi Yang, Marianne D. Keller, Richard Beare, et al. 2011. "Segmentation of the Mouse Hippocampal Formation in Magnetic Resonance Images." *NeuroImage* 58 (3): 732–40. <https://doi.org/10.1016/j.neuroimage.2011.06.025>.
- Rosa, Marcello G.P, and Leah A Krubitzer. 1999. "The Evolution of Visual Cortex: Where Is V2?" *Trends in Neurosciences* 22 (6): 242–48. [https://doi.org/10.1016/S0166-2236\(99\)01398-3](https://doi.org/10.1016/S0166-2236(99)01398-3).
- Steadman, Patrick E., Jacob Ellegood, Kamila U. Szulc, Daniel H. Turnbull, Alexandra L. Joyner, R. Mark Henkelman, and Jason P. Lerch. 2014. "Genetic Effects on Cerebellar Structure Across Mouse Models of Autism Using a Magnetic Resonance Imaging Atlas: MRI of Genetic Mouse Model's Cerebellum." *Autism Research* 7 (1): 124–37. <https://doi.org/10.1002/aur.1344>.
- Strand, Andrew D, Aaron K Aragaki, Zachary C Baquet, Angela Hodges, Philip Cunningham, Peter Holmans, Kevin R Jones, Lesley Jones, Charles Kooperberg, and James M Olson. 2007. "Conservation of Regional Gene Expression in Mouse and Human Brain." Edited by Jonathan Flint. *PLoS Genetics* 3 (4): e59. <https://doi.org/10.1371/journal.pgen.0030059>.
- Ullmann, Jeremy F.P., Charles Watson, Andrew L. Janke, Nyoman D. Kurniawan, and David C. Reutens. 2013. "A Segmentation Protocol and MRI Atlas of the C57BL/6J Mouse Neocortex." *NeuroImage* 78 (September): 196–203. <https://doi.org/10.1016/j.neuroimage.2013.04.008>.
- Zilles, Karl. 2018. "Brodmann: A Pioneer of Human Brain Mapping—His Impact on Concepts of Cortical Organization." *Brain* 141 (11): 3262–78. <https://doi.org/10.1093/brain/awy273>.