

# Transcriptomic similarity in the mouse and human brain

*Antoine Beauchamp*

*December 17, 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, Mossmann, 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.

## 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 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

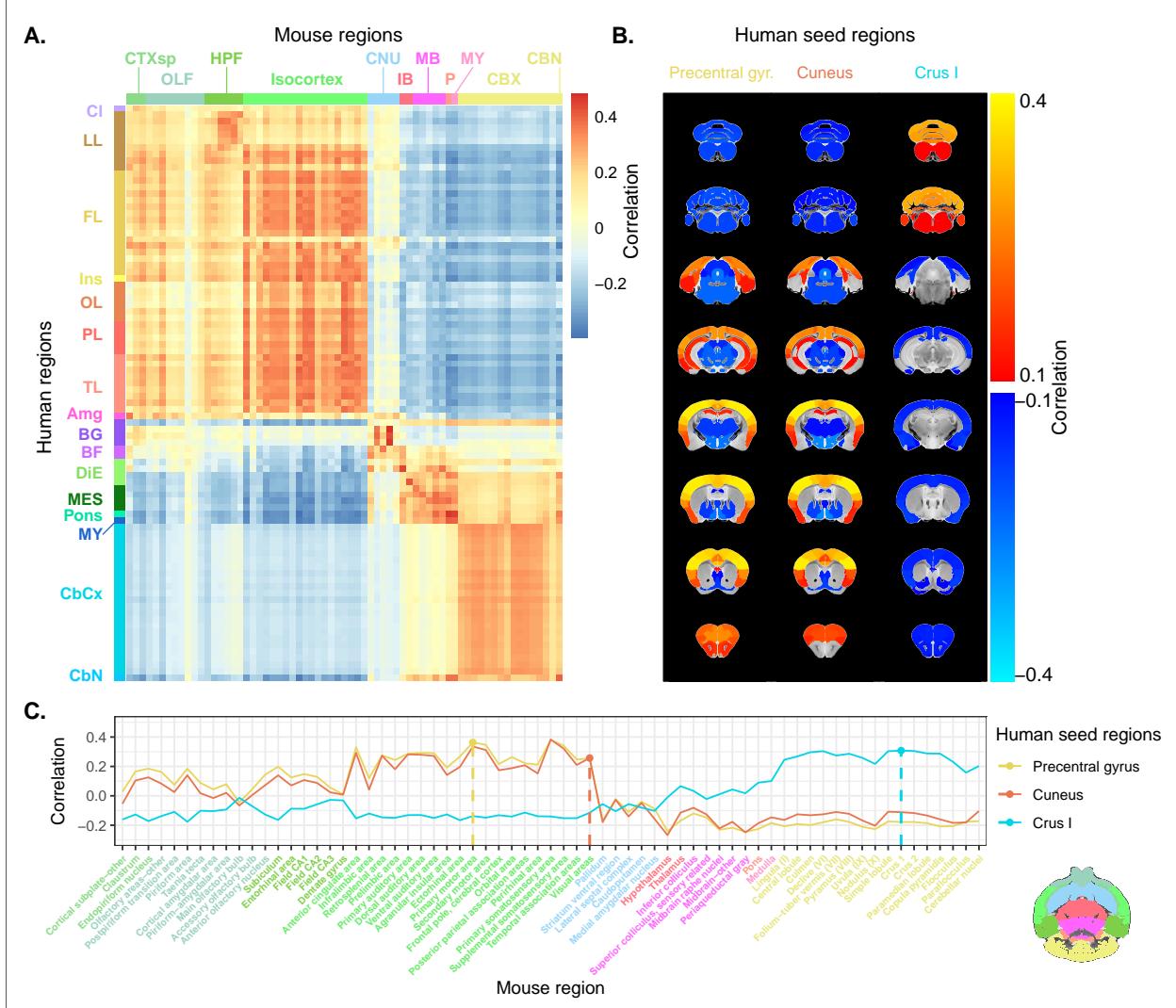


Figure 1: Transcriptomic similarity in the mouse and human brains. (A) Similarity matrix displaying the correlation between 67 mouse regions and 88 human regions based on the expression of 2624 homologous genes. Columns annotated with 11 broad mouse regions: Cortical subplate (CTXsp), olfactory areas (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.

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.

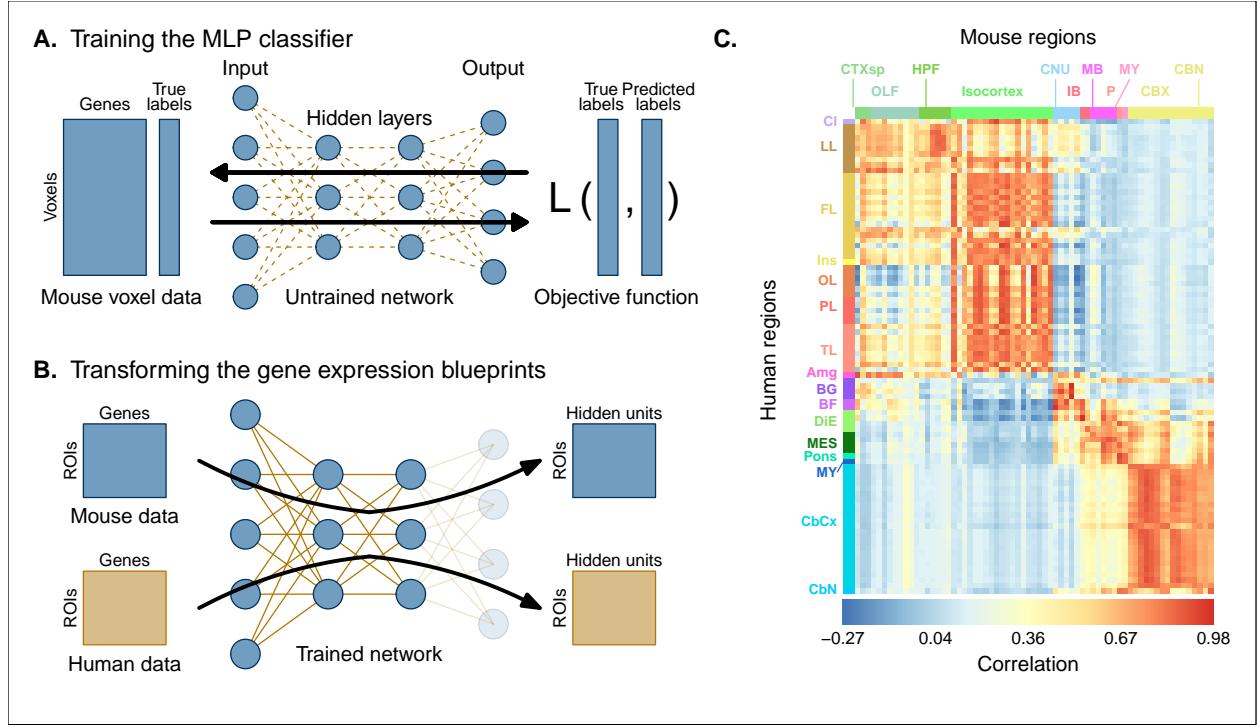


Figure 2: Transforming the common space. (A) Schematic of the neural network training process. Voxel-wise expression maps from 2624 homologous genes in the AMBA were used to train the neural network to classify each mouse voxel into one of 67 atlas regions. (B) Schematic illustrating the transformation 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. 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.

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 ( $\mu = 9.1$ ;  $\sigma = 8.5$ ). 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.

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

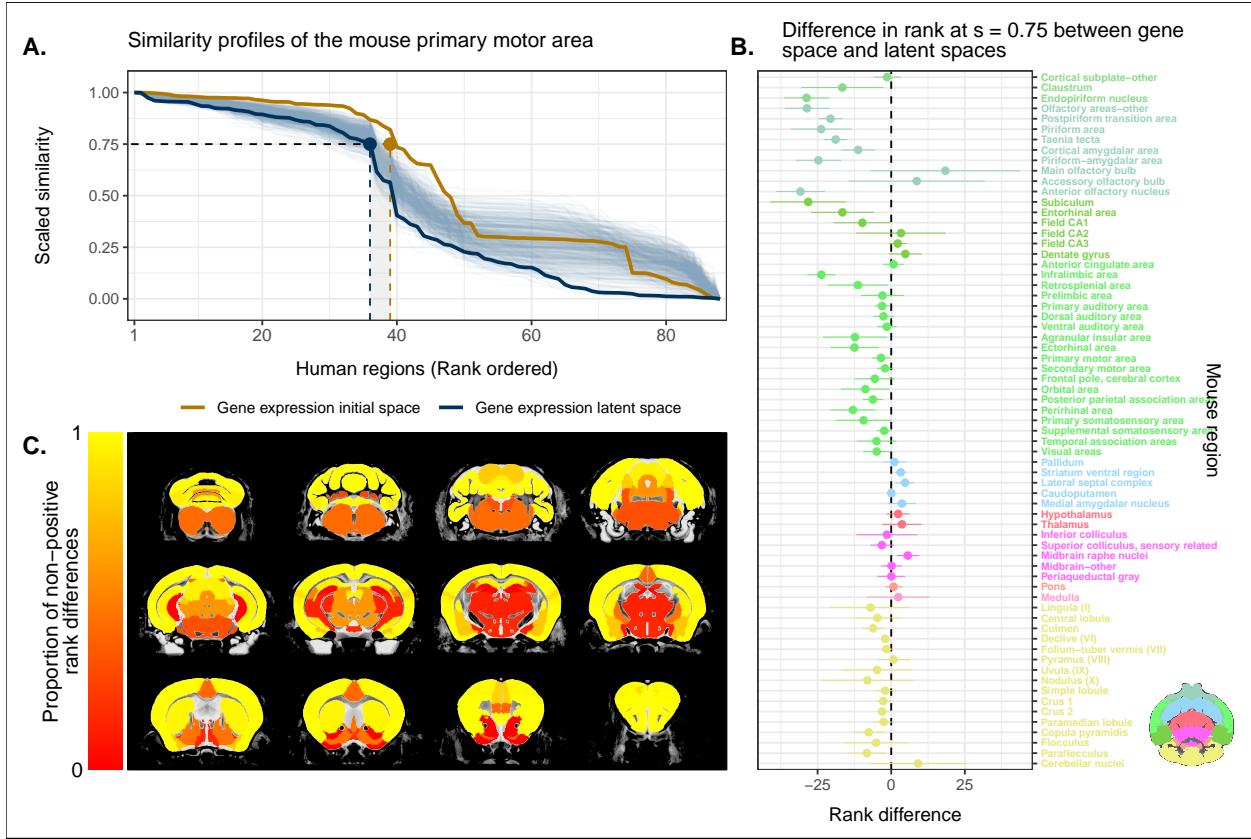


Figure 3: Quantifying improvement in locality in gene expression latent space. (A) The amount of local signal within a broadly similar region of the brain for a finer seed region's (e.g. primary motor area) similarity profile can be quantified by the decay rate of the head of the rank-ordered profile. 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.

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.

## Mouse-human isocortical similarity features sensorimotor and supramodal clusters

Brain regions in the mouse and human isocortices were among those that saw improvement in the gene expression latent spaces, both in terms of locality of matches and similarity of well-established neuroanatomical homologues. It is well-known that gene expression in the cortex codes primarily for different cortical layers rather than areal delineations (Lein et al. 2007). While this is true, our results demonstrate that mouse-human homologous genes also contain information about cortical areas. This is in line with the findings

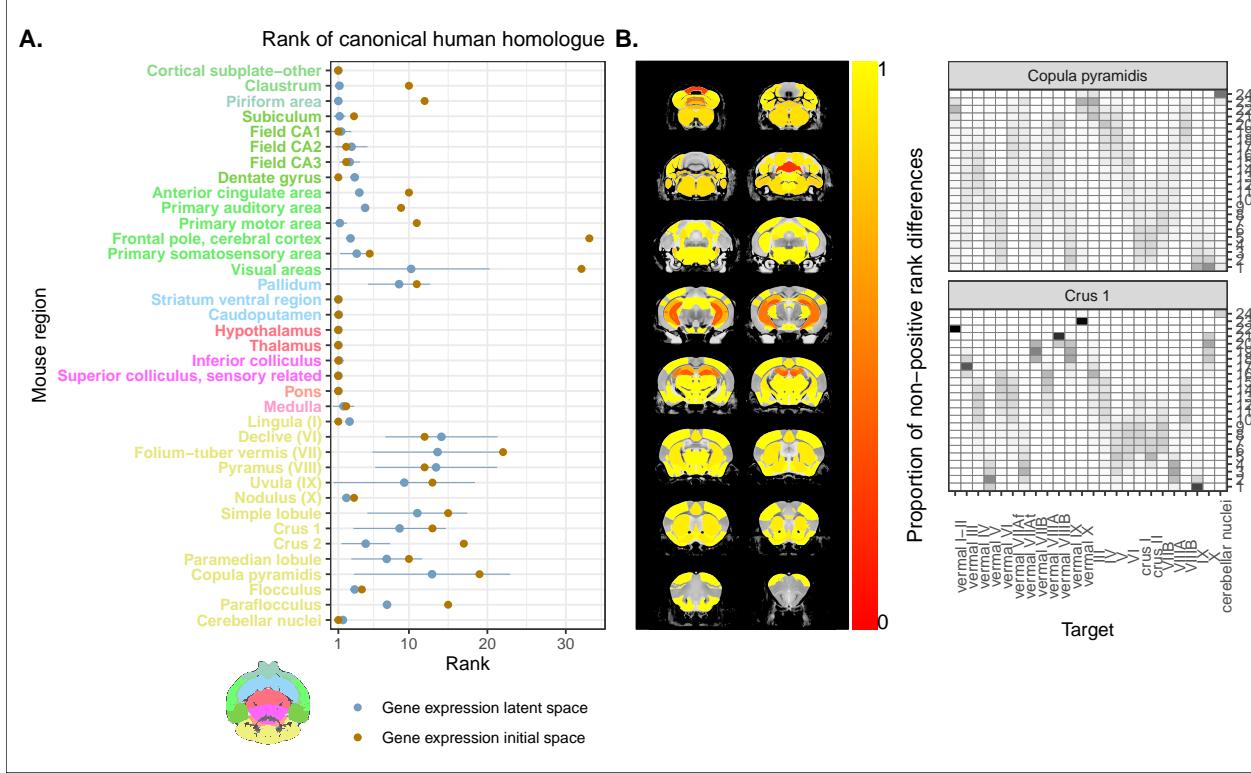


Figure 4: Recovering canonical neuroanatomical pairs in gene expression space. (A) Comparison between the ranks of canonical human matches for mouse seed regions between the initial gene expression space and gene expression latent spaces. Points and error bars represent mean and 95% confidence interval. 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)

reported in a recent study by Ortiz et al. (Ortiz et al. 2020), in which they generated a molecular atlas of the mouse brain using spatial transcriptomic data for 7663 genes. Their parcellation of the isocortex into 55 molecular clusters contains domains pertaining to different cortical layers, but also features segregation of cortical areas along anterior-posterior and medial-lateral axes. Finding improvements in our ability to resolve mouse-human cortical matches, we were curious about the pattern of similarity within the isocortices of the two species. To investigate this, we used hierarchical clustering to cluster mouse and human isocortical regions on the basis of their correlation values, averaged over the gene expression latent spaces (Figure 5, panel A). We loosely identified four main clusters by visually examining the “elbow plots” for the mouse and human clusters (Figure 5, panel B). We find a striking segregation of the mouse isocortex into three main clusters that correspond to regions that are primarily engaged in sensorimotor processing and regions that are supramodal. While all of the sensorimotor areas cluster together in the three-cluster solution, three supramodal areas also form part of this cluster: the retrosplenial area, the posterior parietal association areas, and the anterior cingulate cortex. Of these, the retrosplenial area is the most different, belonging to its own cluster in the four-cluster solution, and in fact is the mouse isocortical region with the smallest degree of similarity to the human cortex. The story is not as clean on the human side. Under the four-cluster solution, the majority of areas form a large cluster that includes a mix of sensorimotor and supramodal regions. However if the number of clusters is increased to 5, this large cluster breaks up into two smaller clusters that feature some delineation between supramodal and sensorimotor areas, which are primarily motor and auditory in nature (e.g. precentral gyrus, Heschl’s gyrus). Interestingly, the postcentral gyrus (i.e. primary somatosensory area) forms a separate cluster with a host of visual areas like the cuneus and lingual gyrus. These regions exhibit very similar correlation profiles to the mouse isocortical regions, including maximal correlation to the mouse primary somatosensory area, with an average of  $r = 0.92$ . Thus the human sensorimotor isocortical regions are loosely clustered according to sensory-visual areas and auditory-motor areas (**Association/supramodal and sensorimotor ROI annotation on the human side as well?**). We additionally ran hierarchical clustering on the isocortical similarity matrix in the original homologous gene space. While the cluster annotations were not substantially different (**Might need supplemental figure for this**), we observed that the Euclidean distances within and between clusters were smaller compared with the latent space clustering (Figure 5, panel B). This further confirms that the MLP classification approach improves the segregation of brain regions in gene space.

Mouse isocortical areas clustered nicely along the lines of sensorimotor and supramodal functional types based on their similarity to human cerebral regions. We explored whether this categorization was associated with differences in correlation strength to the human brain. To evaluate this, we calculated the maximal correlation value for every region in the mouse isocortex in each of the 500 gene expression latent spaces (Figure 5, panel C). We also computed the average of these maximal correlations for the sensorimotor and supramodal regions (Figure 5, panel D). On average we find that sensorimotor regions are associated with higher maximal correlation values over the latent spaces ( $r = 0.89 \pm 0.04$ ) compared with supramodal areas ( $r = 0.85 \pm 0.03$ ). This is in line with our expectation that regions involved in more basic sensory and motor processing should exhibit a stronger degree of similarity between mouse and human brains. We find that the mouse primary somatosensory and motor areas have the highest average maximal correlation values, with  $r = 0.94 \pm 0.04$  and  $r = 0.93 \pm 0.04$  respectively. These maximal values provide a sense of the strength of association between regions, but don’t speak to the uniqueness of the matches. For instance, as we observed in the cluster analysis, the primary somatosensory area exhibits high correlation values to the postcentral gyrus but also to five additional regions in the occipital lobe. This can also be seen with the primary motor

area: The precentral gyrus is one of the most similar regions in the human brain, but so are the inferior and superior occipital gyri. In contrast, the retrosplenial area and the frontal pole are the mouse isocortical regions that are least similar to regions in the human cerebrum. Interestingly, while the mouse frontal pole is not as strongly correlated to the human regions, it is most strongly associated with the human frontal pole (Figure 5, panel A; Figure 4, panel A).

Overall we find that the expression of homologous genes captures sub-regional organization in the mouse and human isocortices. While the pattern of correspondences is not at the level of individual cortical areas, there are clear distinctions between areas that are sensorimotor and those that are supramodal, especially for regions in the mouse isocortex. Within these clusters, individual cortical areas are highly correlated to a number of regions in the brain of the other species. The majority of isocortical regions exhibit both divergent one-to-many and convergent many-to-one kinds of associations. For example the mouse primary somatosensory area have a diffusive pattern of similarity to the human postcentral gyrus and multiple visual areas, but each of these human regions in turn has a convergent pattern of similarity to the mouse somatosensory cortex, while also being diffusively associated with, for instance, the primary motor area. We also find that, while the identity of the most similar matches may vary, mouse sensorimotor areas are generally more similar than supramodal areas. This is reasonable from an evolutionary perspective. While certain sensorimotor areas have certainly become more specialized in either species over the course of evolution (**need ref**), the basic neurobiological machinery is likely to have been inherited from the last common ancestor rather than developed independently (**ref**). (**Finish this**)

## Transcriptomic similarity reveals complex relationships between the mouse and human striatum

An advantage of the common space approach is that it allows for the quantitative comparison of brain organization across 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 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, 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

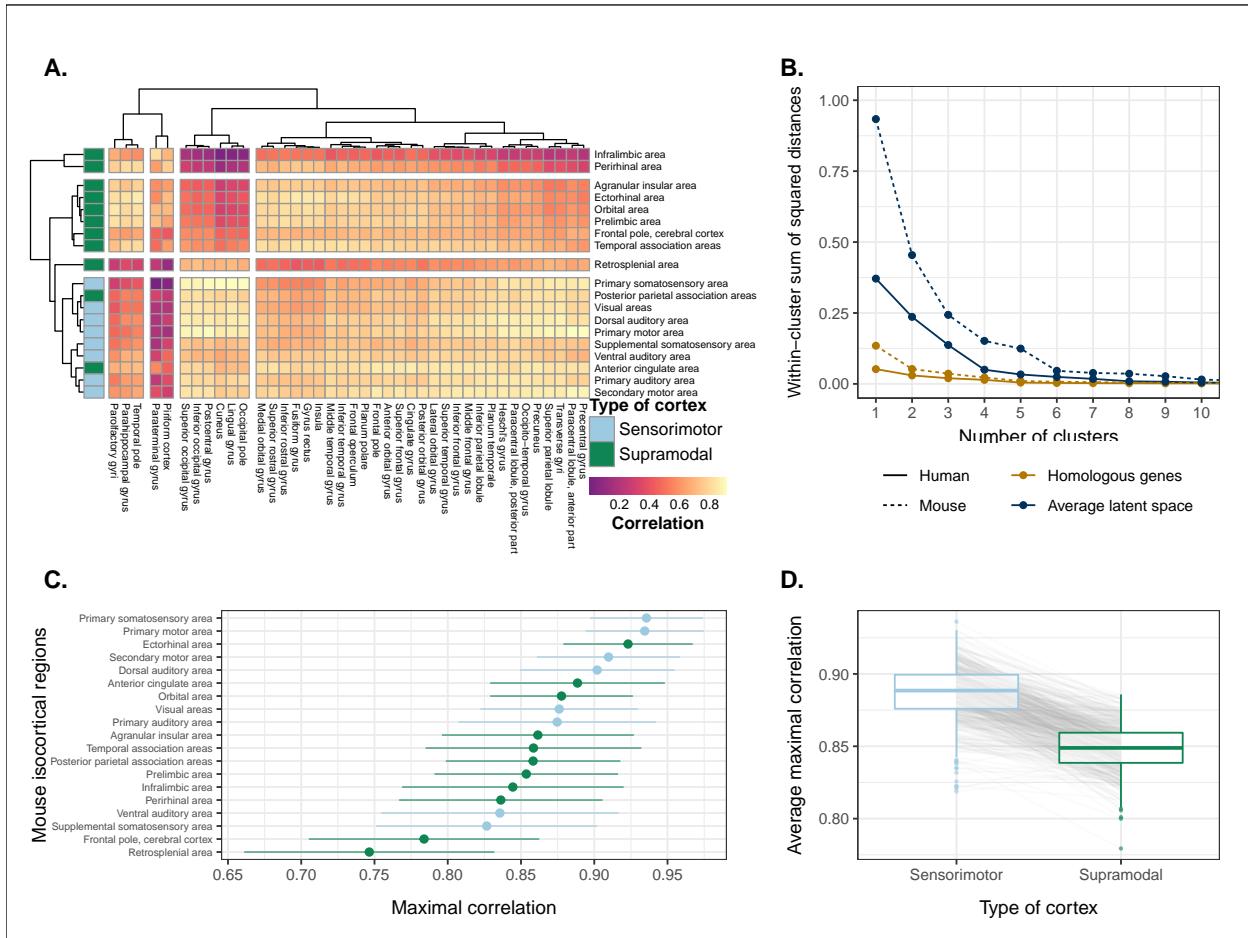


Figure 5: Organization of mouse-human isocortical similarity. (A) Hierarchical clustering of mouse and human isocortical regions based on average latent space correlation values. Mouse regions are annotated as sensorimotor or supramodal. (B) Within-cluster sum of squared distances for different numbers of mouse and human isocortical clusters in the average latent space and initial homologous gene space. We loosely selected an optimal cluster number equal to 4. (C) Maximal correlation values of mouse isocortical region. Points and error bars represent mean and 95% confidence interval over latent space samples. (D) Average maximal correlation per sensorimotor or supramodal isocortical areas in each gene expression latent space. Grey lines correspond to individual latent spaces.

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. The median of the distribution for the caudate-caudoputamen pairs is 0.95 and 0.97 for the putamen-caudoputamen pairs, with modal values of 0.94 and 0.97, respectively. All latent spaces return correlations greater than 0.9 for caudate-caudoputamen and putamen-caudoputamen pairs. Beyond this expected top match, the caudate and putamen both exhibit high similarity to the nucleus accumbens and the fundus of striatum, with mean correlation values of about 0.80. 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 variance pooled over all mouse targets (**Jason comment: Interpretation of pooled variance not obvious**) is much lower ( $\sigma = 0.04$ ) compared with the pooled variances for the caudate ( $\sigma = 0.09$ ) and putamen ( $\sigma = 0.10$ ), indicating less specificity to any one mouse striatal target. In particular, the human nucleus accumbens isn't as specifically similar to the mouse nucleus accumbens in the way that the caudate and putamen are similar to the caudoputamen. The mouse target distributions are right-shifted compared with those for the caudate and putamen, with mean values of 0.90, 0.89, and 0.89 for the mouse nucleus accumbens, caudoputamen, and fundus of striatum, respectively. The human accumbens also exhibits a high degree of similarity to the mouse olfactory tubercle, the distribution of which is also right-shifted compared with the caudate and putamen.

Given the high correlation of the human caudate and putamen to the mouse caudoputamen, as well as the finding reported by Balsters et al. about the similarity of the lateral caudoputamen to the putamen, we were curious as to whether we could identify sub-regional patterns of similarity in the caudoputamen and other striatal regions using these gene expression data. To probe this question, we first examined the average latent space correlation between each voxel in the mouse striatum and every region in the human atlas. We created brain maps for the human regions that exhibited the highest mean correlation values, averaged over mouse striatal voxels: the caudate, the putamen, the nucleus accumbens, and the septal nuclei (Figure 6, panel B). We find that voxels in the caudoputamen exhibit a homogeneous pattern of similarity to both the caudate and the putamen. On average, voxels in the caudoputamen have a correlation of 0.95 to the caudate and 0.94 to the putamen, with standard deviations of 0.04 and 0.05 respectively. The caudate and putamen are associated with correlations of at least 0.90 in 88% and 84% of caudoputamen voxels. A number of voxels are also highly similar to the human nucleus accumbens, with an average correlation value of 0.90 and 55% of voxels returning a correlation of at least 0.9. The caudoputamen voxels most similar to the nucleus accumbens lie in the ventral-rostral part of the region. Of course, voxels in the mouse nucleus accumbens are also highly similar to the human nucleus accumbens, with an average of 0.90 and standard deviation of 0.06. While the human nucleus accumbens is the most strongly correlated region, a number of voxels also exhibit reasonably strong correlations to the substantia innominata, the septal nuclei, and the amygdala. Indeed, 91% of voxels in the accumbens are correlated at a value of 0.7 or higher to the substantia innominata. The equivalent percentages for the septal nuclei and amygdala are 78% and 74% respectively.

We additionally examined the proportion of latent spaces in which each voxel in the mouse striatum was maximally similar to the human target regions (Figure 6, panel C). As expected, we find that voxels in the caudoputamen are most often maximally similar to the human caudate and putamen, with 75% of voxels in the caudoputamen being maximally similar to the caudate or putamen in at least 95% of latent spaces, and 62% of voxels being maximally similar to one of those targets in *all* latent spaces. Interestingly, we observe the emergence of a continuous bilateral pattern of specificity to the caudate and putamen, with voxels

in the rostral and lateral-caudal parts of the caudoputamen being more specific to the caudate, and voxels in the medial-rostral part being more specific to the putamen. This map highlights subtle differences in the similarity between caudoputamen voxels and the caudate or putamen. While this pattern distinguishes the two regions on the basis of which is the top match, individual voxels have very similar correlation values to the targets (Figure 6, panel B), with a mean difference in correlation of only 0.006. Beyond the caudoputamen, we find that the accumbens and olfactory tubercle in the mouse are consistently similar to the human nucleus accumbens, with 80% of mouse accumbens voxels and 65% of olfactory tubercle voxels having the human accumbens as their top match in at least 80% of latent spaces. For those voxels below this threshold, the human regions that are most often the top match are once again the amygdala, the septal nuclei, and the substantia innominata.

Overall, we observe a strong association between the mouse caudoputamen and both the human caudate and putamen. While we find a subtle pattern of specificity to either region among voxels in the caudoputamen on the basis of maximal similarity, the high degree of similarity in the correlation values to each region suggests that the majority of voxels in the caudoputamen are equally similar to the caudate and the putamen on the basis of the expression of mouse-human homologous genes. This is in contrast to the findings reported by Balsters et al., in which which in the human posterior putamen were statistically similar to the lateral portion of the caudoputamen. We also find that the nucleus accumbens is well conserved across species. However the region also exhibits patterns of similarity that go beyond the simple one-to-one match. The human accumbens features similar correlation values to the mouse caudoputamen and fundus of striatum, in addition to the accumbens proper, with no sharp distinction between these regions. It also exhibits a larger degree of similarity to the mouse olfactory tubercle. This is also seen in the mouse striatum, where voxels in the accumbens and the olfactory tubercle map onto the human accumbens.

## 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. Using the expression of 2624 homologous genes, we identified broad patterns of similarity between the brains of the two species. 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 the lower-dimensional data representations, we demonstrated that mouse and human isocortical regions segregate broadly into sensorimotor and supramodal clusters, with this pattern being more well-defined for mouse areas. We also showed that mouse sensorimotor areas are more strongly correlated to human regions on average, compared with supramodal areas. Finally, we explored patterns of similarity in the mouse and human striata. We found that the human caudate and putamen were highly correlated with the mouse caudoputamen, and vice versa, with no evident sub-regional pattern of specificity within the caudoputamen to either of the human regions. We also found a strong degree of similarity between the nucleus accumbens in both species. However, the human nucleus accumbens was additionally similar to the caudoputamen, fundus of striatum and olfactory tubercle. Voxels in the mouse accumbens were primarily similar to the human accumbens but also featured patterns of similarity to the substantia innominata and amygdala.

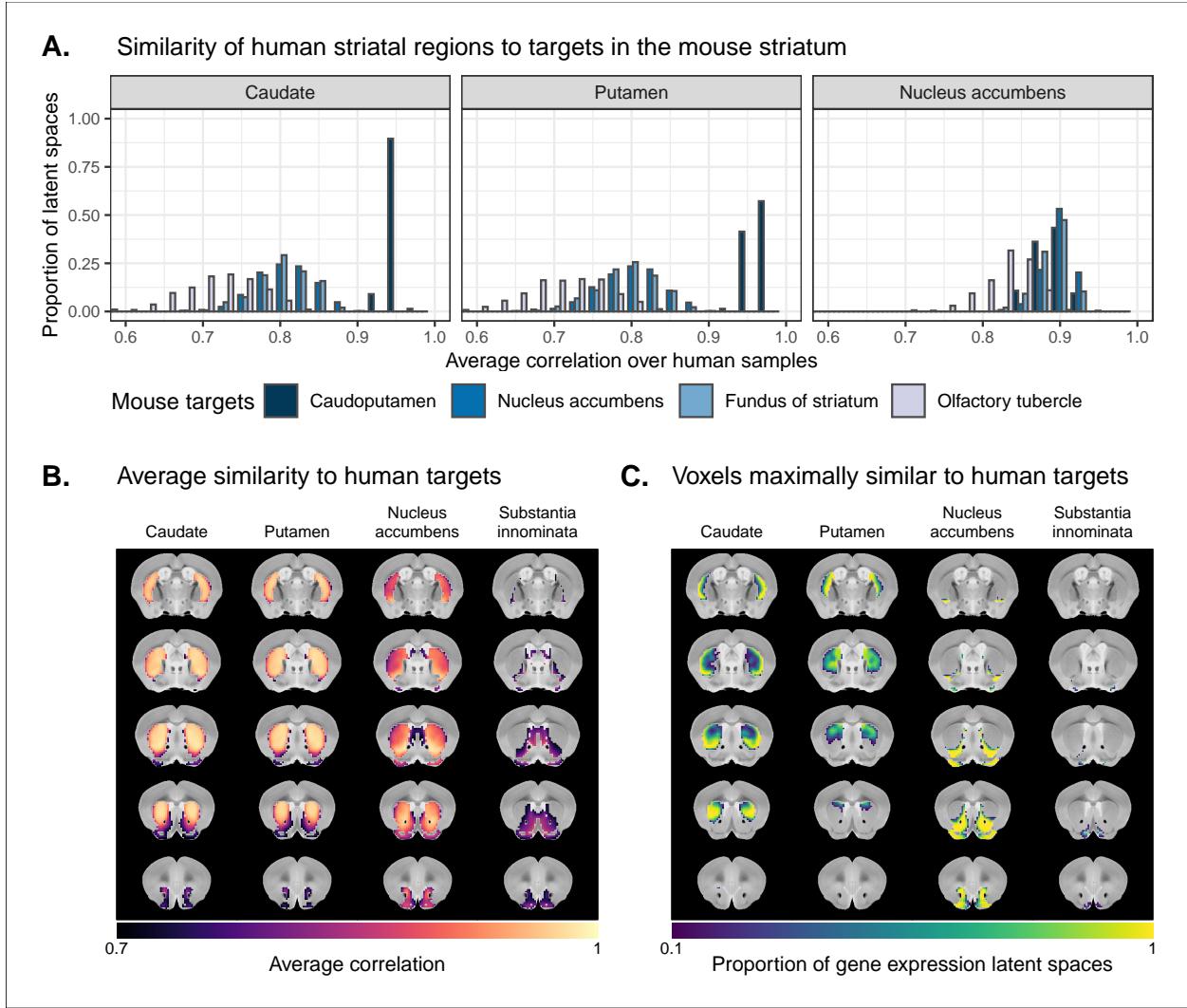


Figure 6: Similarity among mouse and human striatal regions. (A) Distributions over gene expression latent spaces of region-wise average correlation values for mouse and human striatal pairs. Human regions were chosen based on the AHBA ontology. Mouse target regions were chosen to be those with the highest average correlation values. (B) Latent space averaged correlations between voxels in the mouse striatum and human target regions. Target regions were selected based on the highest mean correlation across all striatal voxels. (C) 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 (Horev et al. 2011; Gompers et al. 2017; Pagani et al. 2021). It is common for researchers involved in translational neuroscience to rely on findings of this kind to make inferences about the human disorder. The typical approach, which is to identify rough post-hoc correspondences between neuroanatomical ontologies, is not particularly comprehensive and is subject to confirmation bias. While it may be 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 (Kaas 2005; Krubitzer and Kaas 2005). 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 qualitative one-to-one approach does not capture the entire story. A prime example of this is the separation of the caudoputamen into anatomically distinct nuclei, the caudate and the putamen (**Not sure where I'm going with this.**). Novel quantitative methods, such as the common space approach, are needed to improve the robustness of cross-species comparisons and to ameliorate the evaluation of face validity for animal models used in translational research.

The expression of homologous genes provides an elegant way to define a common space for quantitative cross-species comparisons since it relies on homology at a deep molecular biological level. The approach is not without limitations however. The acquisition of whole-brain 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 types of analyses that are possible (e.g. null-hypothesis significance testing) and largely limits the availability of replication data sets. That being said, new technologies, such as spatial transcriptomics, are gradually making it easier to acquire brain-wide gene expression data in less time and at lower cost (Stahl et al. 2016; Vickovic et al. 2019; Ortiz et al. 2020). In the meantime, other modalities such as rs-fMRI have an advantage. Larger samples can be generated with relative ease in a variety of species. The caveat here is that the resulting common space relies on a definition of homology at a higher level, i.e. a priori definitions of homologous neuroanatomical targets. Realistically, both of these modalities are useful in their own right and likely provide complementary information about mouse-human brain correspondences. For instance, we have shown that neuroanatomical organization within the cerebellar cortex is poorly captured by the expression of mouse-human homologous genes. This is not limited to cross-species comparisons: It has been shown both in the mouse and the human separately that gene expression similarity within the cerebellar cortex is incredibly homogeneous (Lein et al. 2007; Hawrylycz et al. 2012; Myers 2017). Thus transcriptomic data is not well-suited to investigating cerebellar organization. In contrast, modalities such as functional connectivity and structural covariance have proven useful for elucidating sub-regional organization in the cerebellum (Yee et al. 2018; Ren, Guo, and Guo 2019). 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 (Wang et al. 2014) to

generate a multi-modal similarity network between mouse and human brain regions or voxels. Strategies like this one may work better down the road, as any one modality may not be well-suited to the investigation of specific brain regions.

The common space framework is a powerful way to explore and evaluate relationships between the brains of different organisms. 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, the framework allows for the quantification of cross-species comparisons, which is needed to reduce bias in interpreting translational results. Additionally, common spaces open the door to new ways of translating results between species, such as the direct transformation of brain-wide maps from one species to the other (Rogier B Mars et al. 2018; Mandino et al. 2021). (**Closing statement**)

## Materials and methods

### Mouse gene expression data

We used the whole-brain in-situ hybridization (ISH) coronal data set from the Allen Mouse Brain Atlas (Lein et al. 2007). Specifically, we used gridded gene expression data, i.e. expression data aligned to the Allen Mouse Brain Common Coordinate Framework (CCFv3)(Wang et al. 2020) and summarized under a lower resolution grid ( $200\mu m$ ). We used gene expression “energy” values, which are defined by the Allen Institute at each voxel as the product of the expression density (“sum of expressing pixels / sum of all pixels in division”) and expression intensity (“sum of expressing pixel intensity / sum of expressing pixels”) (see <http://help.brain-map.org/display/mousebrain/API> for more details). Mouse gene expression data were obtained from the Allen Institute’s API ([http://api.brain-map.org/grid\\_data/download/%5Bgene id\]](http://api.brain-map.org/grid_data/download/%5Bgene id]), see <https://help.brain-map.org/display/api/Downloading+3-D+Expression+Grid+Data> for more details) as a raw sequence of float (32-bit) values and subsequently reshaped into 3D images in the MINC format using the `rawtominc` utility. Origin, extents, and spacing were defined such that the image was RAS-oriented, with the origin at the point where the anterior commissure crosses the midline. The MINC images for 4346 ISH experiments were imported into the Python programming language using the `pyminc` library ([ref?](#)). The image arrays were masked and reshaped to form an experiment-by-voxel expression matrix. We pre-processed this data by first applying a `log2` transformation for consistency with the human data set. For those genes associated with more than one ISH experiment, we averaged the expression of each voxel across the experiments. We subsequently filtered out genes for which more than 20% of voxels contained missing values. Finally we applied a K-nearest neighbours algorithm to impute the remaining missing values. This was done using the genes as the variables and voxels as the observations. The result of this pre-processing pipeline 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 six donors, each of which contains  $\log_2$  expression values for 58692 gene probes across numerous tissue samples. The data were pre-processed using a custom pipeline built following the recommendations from Arnatkeviciūtė et al. (Arnatkeviciūtė, Fulcher, and Fornito 2019). The pipeline was implemented using the R programming language. Specifically, once imported, we passed the data from individual donors 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 six 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 from the Mouse Imaging Centre (Dorr et al. 2008; Richards et al. 2011; Ullmann et al. 2013; Steadman et al. 2014; Qiu et al. 2018). The labels of the DSURQE atlas correspond to the leaf node regions in the AMBA ontology, which allowed us to use the hierarchical neuroanatomical tree to aggregate and prune the atlas labels to the desired level of granularity. For the purposes of our analyses, we removed white matter and ventricular regions entirely. The remaining grey matter regions were aggregated up the hierarchy so that the majority of resulting labels contained enough voxels to be classified appropriately by the multi-layer perceptron. In doing so, we maintained approximately the same level of tree depth within a broad region (e.g. cerebellar regions were chosen at the same level of granularity). This resulted 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 aggregated and pruned the neuroanatomical hierarchy to correspond roughly to the level of granularity obtained 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.

## Expression and similarity matrices

We created the mouse and human gene-by-region expression matrices from the human gene-by-sample expression matrix and the mouse gene-by-voxel expression matrix. We intersected these gene sets with a list of 3331 homologous genes obtained from the NCBI HomoloGene database (NCBI 2018) (**Yohan: What is**

**considered a homologous gene?**), resulting in 2624 genes present in both the mouse and human data sets. We then annotated each of the human samples with one of the 88 human atlas regions, and each of the mouse voxels with one of the 67 mouse atlas regions, discarding white matter and ventricular entries in the process. These labelled expression matrices were subsequently normalized. For each matrix, we first standardized every gene across all voxels or samples using a z-scoring procedure. We then centered every voxel or sample by subtracting the average expression over all homologous genes. Finally, we generated the gene-by-region matrices by averaging the expression of every gene over the voxels or samples corresponding to each atlas region. Using these expression matrices, we generated the mouse-human similarity matrix by computing the Pearson correlation coefficient between all pairs of mouse and human regions.

## Multi-layer perceptron classification and transformation

To improve the resolution of mouse-human neuroanatomical matches, we performed a supervised learning approach, wherein we trained a MLP neural network to classify 67 mouse atlas regions from the expression values of 2624 homologous genes. We chose a model architecture in which each layer of the network was fully connected to previous and subsequent layers. To optimize the hyperparameters, we implemented an ad hoc cross-validation procedure that took into account the fact that the majority of genes in the coronal AMBA data set are sampled only once over the entire mouse brain. The procedure involved a combination of the coronal data set and the sagittal AMBA data set. The genes from the sagittal experiments were filtered to match the genes in the coronal set. We imported and processed both the sagittal and coronal data sets using a version of the pipeline described above with the following modifications: 1. We applied a *unilateral* brain mask to both the coronal and sagittal data sets since the sagittal data is unilateral by construction, and 2. we did not aggregate the expression of multiple ISH experiments for those genes in the coronal set featuring more than one experiment. Once these experiment-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 also annotated the voxels in each of the expression matrices with one of the 67 regions in the atlas. Our validation procedure then involved iterative construction of training and validation sets by sampling gene experiments from either the coronal or sagittal data sets: For every gene in the homologous set, we first determined whether that gene was associated with more than one experiment in the coronal 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. If the gene was associated with only one experiment in the coronal set, we randomly chose either the coronal or sagittal experiment for the training set and the other for the validation set. Once the training and validation sets were generated, they were normalized as described above. We then trained the MLP using the training set and evaluated its performance on the validation set. Given that the construction of the training and validation sets involved some randomness, we repeated this construction, training, and validation procedure 10 times for every hyperparameter combination.

The hyperparameters that we optimized using this method were 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 weight decay. The values we optimized over were as follows. We varied the number of hidden layers between 3 and 5. We varied the number of hidden units between 100 and 1000 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  $10^{-5}$ . 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 returned an average classification

accuracy of 0.926 on the training sets and of 0.526 on the validation sets. Following validation, we used the optimal hyperparameters to train the MLP on the full bilateral coronal voxel-wise expression matrix. These models were implemented in Python using `NeuralNetClassifier` class from the `skorch` API (see <https://skorch.readthedocs.io/en/stable/#>). For both validation and training, the models were trained over 200 epochs using a maximum learning rate of  $10^{-5}$ . We used the `AdamW` optimization algorithm (Loshchilov and Hutter 2019) and `OneCycleLR` learning rate scheduler policy from PyTorch. The activation function used in the forward pass was the rectified linear unit (ReLU) and the loss function was the negative log-likelihood loss, which is the default for the `NeuralNetClassifier` class.

We used the trained MLP to generate the new gene expression latent space. To extract the appropriate transformation, we removed the predictive output layer and softmax transformation from the network architecture. The resulting architecture returns the 200 hidden units in the third hidden layer as the output of the network. To create the latent space data representations, we applied this network to the mouse and human gene-by-region expression matrices, transposed so that the genes were the input variables. The resulting matrices have 200 columns corresponding to the hidden units and 67 (88) rows corresponding to the mouse (human) regions. They describe each brain region's weight over the hidden units. To create the similarity matrix, we calculated the Pearson correlation coefficient between all pairs of mouse and human regions.

Given the stochasticity inherent in training the MLP (e.g. random weight initialization and stochastic optimization), we repeated the training and transformation process 500 times using the same network architecture and input data.

## References

- Arnatkeviciūtė, 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>.
- Bronchi, 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>.
- Gompers, Andrea L, Linda Su-Feher, Jacob Ellegood, Nycole A Copping, M Asrafuzzaman Riyadh, Tyler W Stradleigh, Michael C Pride, et al. 2017. “Germline Chd8 Haploinsufficiency Alters Brain Development in Mouse.” *Nature Neuroscience* 20 (8): 1062–73. <https://doi.org/10.1038/nn.4592>.

- 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.
- Horev, G., J. Ellegood, J. P. Lerch, Y.-E. E. Son, L. Muthuswamy, H. Vogel, A. M. Krieger, et al. 2011. “Dosage-Dependent Phenotypes in Models of 16p11.2 Lesions Found in Autism.” *Proceedings of the National Academy of Sciences* 108 (41): 17076–81. <https://doi.org/10.1073/pnas.1114042108>.
- 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>.
- Loshchilov, Ilya, and Frank Hutter. 2019. “Decoupled Weight Decay Regularization.” In *Proceedings of the Seventh International Conference on Learning Representations*. New Orleans. <http://arxiv.org/abs/1711.05101>.
- Mandino, Francesca, Roël M. Vroooman, Heidi E. Foo, Ling Yun Yeow, Thomas A. W. Bolton, Piergiorgio Salvan, Chai Lean Teoh, et al. 2021. “A Triple-Network Organization for the Mouse Brain.” *Molecular Psychiatry*, October, 1–8. <https://doi.org/10.1038/s41380-021-01298-5>.
- 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 Sotiropoulos, 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.
- Pagani, Marco, Noemi Barsotti, Alice Bertero, Stavros Trakoshis, Laura Ulysse, Andrea Locarno, Ieva Miseviciute, et al. 2021. "mTOR-Related Synaptic Pathology Causes Autism Spectrum Disorder-Associated Functional Hyperconnectivity." *Nature Communications* 12 (1): 6084. <https://doi.org/10.1038/s41467-021-26131-z>.
- 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>.
- Ren, Yudan, Lei Guo, and Christine Cong Guo. 2019. "A Connectivity-Based Parcellation Improved Functional Representation of the Human Cerebellum." *Scientific Reports* 9 (June): 9115. <https://doi.org/10.1038/s41598-019-45670-6>.
- 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).
- Stahl, Patrick L., Fredrik Salmén, Sanja Vickovic, Anna Lundmark, José Fernández Navarro, Jens Magnusson, Stefania Giacomello, et al. 2016. "Visualization and Analysis of Gene Expression in Tissue Sections by Spatial Transcriptomics." *Science* 353 (6294): 78–82. <https://doi.org/10.1126/science.aaf2403>.
- 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 Hollmans, 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>.

Vickovic, Sanja, Gökcen Eraslan, Fredrik Salmén, Johanna Klughammer, Linnea Stenbeck, Denis Schapiro, Tarmo Äijö, et al. 2019. “High-Definition Spatial Transcriptomics for in Situ Tissue Profiling.” *Nature Methods* 16 (10): 987–90. <https://doi.org/10.1038/s41592-019-0548-y>.

Wang, Bo, Aziz M Mezlini, Feyyaz Demir, Marc Fiume, Zhuowen Tu, Michael Brudno, Benjamin Haibe-Kains, and Anna Goldenberg. 2014. “Similarity Network Fusion for Aggregating Data Types on a Genomic Scale.” *Nature Methods* 11 (3): 333–37. <https://doi.org/10.1038/nmeth.2810>.

Wang, Quanxin, Song-Lin Ding, Yang Li, Josh Royall, David Feng, Phil Lesnar, Nile Graddis, et al. 2020. “The Allen Mouse Brain Common Coordinate Framework: A 3D Reference Atlas.” *Cell* 181 (4): 936–953.e20. <https://doi.org/10.1016/j.cell.2020.04.007>.

Yee, Yohan, Darren J. Fernandes, Leon French, Jacob Ellegood, Lindsay S. Cahill, Dulcie A. Vousden, Leigh Spencer Noakes, et al. 2018. “Structural Covariance of Brain Region Volumes Is Associated with Both Structural Connectivity and Transcriptomic Similarity.” *NeuroImage* 179 (October): 357–72. <https://doi.org/10.1016/j.neuroimage.2018.05.028>.

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>.