Methods 73 (2015) 79–89

Contents lists available at ScienceDirect

Methods

journal homepage: www.elsevier.com/locate/ymeth

Visualizing the spatial gene expression organization in the brain through non-linear similarity embeddings

Ahmed Mahfouz [a,b,1], Martijn van de Giessen [a,b,1], Laurens van der Maaten [a,b], Sjoerd Huisman [a,b], Marcel Reinders [b], Michael J. Hawrylycz [c], Boudewijn P.F. Lelieveldt [a,b,*]

[a] Division of Image Processing, Department of Radiology, Leiden University Medical Center, Leiden, The Netherlands
[b] Department of Intelligent Systems, Delft University of Technology, Delft, The Netherlands
[c] Allen Institute for Brain Science, Seattle, WA, USA

ARTICLE INFO

Article history:
Received 7 July 2014
Received in revised form 18 September 2014
Accepted 5 October 2014
Available online 16 October 2014

Keywords:
Allen Brain Atlas
Spatial-mapped gene expression
Brain transcriptome structure
Dimensionality reduction
Stochastic Neighbor Embedding

ABSTRACT

The Allen Brain Atlases enable the study of spatially resolved, genome-wide gene expression patterns across the mammalian brain. Several explorative studies have applied linear dimensionality reduction methods such as Principal Component Analysis (PCA) and classical Multi-Dimensional Scaling (cMDS) to gain insight into the spatial organization of these expression patterns. In this paper, we describe a non-linear embedding technique called Barnes-Hut Stochastic Neighbor Embedding (BH-SNE) that emphasizes the local similarity structure of high-dimensional data points. By applying BH-SNE to the gene expression data from the Allen Brain Atlases, we demonstrate the consistency of the 2D, non-linear embedding of the sagittal and coronal mouse brain atlases, and across 6 human brains. In addition, we quantitatively show that BH-SNE maps are superior in their separation of neuroanatomical regions in comparison to PCA and cMDS. Finally, we assess the effect of higher-order principal components on the global structure of the BH-SNE similarity maps. Based on our observations, we conclude that BH-SNE maps with or without prior dimensionality reduction (based on PCA) provide comprehensive and intuitive insights in both the local and global spatial transcriptome structure of the human and mouse Allen Brain Atlases.

1. Introduction

The mammalian brain is a complex system governing all high-level cognitive tasks. The complexity of this system is reflected in the large number of cell types, organized into hundreds of distinct structures [1]. A major challenge facing the neuroscience community is to collect, integrate and analyze data across different levels and scales to produce new insights about the brain's anatomical and functional organization [2]. At the molecular level, each brain structure has a specific cellular composition with a distinct gene expression signature that dictates its functional role [3]. Therefore, to understand the basic anatomical and functional organization of the brain in relation to gene functions, it is crucial to study the spatial localization of genome-wide gene expressions in the brain.

Given the high cellular diversity in the brain, mapping genes at a sufficient spatial resolution is essential to analyze the transcriptome architecture of the brain. Several studies have previously mapped the expression of genes across the mammalian brain, but they have all been limited either in terms of the number of genes analyzed and/or the number of brain structures assessed [4,5]. The Allen Institute for Brain Sciences provides comprehensive genome-wide maps of gene expression across the mouse and human brain, providing a unique opportunity to study the transcriptome architecture of the mammalian brain. In the Mouse Brain Atlas [6] the expression of ~20,000 genes at a cellular resolution using *in situ* hybridization (ISH) is mapped on an anatomical atlas of the mouse brain. Comparably, the Human Brain Atlas [7] employed microarrays to produce a genome-wide map of transcript distribution across the entire human brain. These two resources allow the unprecedented study of how the transcriptome architecture of different brain regions instructs their functional role.

The high diversity of spatially-mapped gene expression patterns in the brain, ranging from globally-expressed genes to

Abbreviations: PCA, Principal Component Analysis; cMDS, classical Multi-Dimensional Scaling; BH-SNE, Barnes-Hut Stochastic Neighbor Embedding; t-SNE, t-Distributed Stochastic Neighbor Embedding; ISH, *in situ* hybridization; API, application programming interface; MR, magnetic resonance; T1W, T1-weighted; T2W, T2-weighted; DT, Diffusion Tensor; AAL, Automated Anatomical Labeling.

* Corresponding author at: Leiden University Medical Center, Division of Image Processing (C2-S), Albinusdreef 2, 2333 ZA Leiden, The Netherlands.
E-mail address: b.p.f.lelieveldt@lumc.nl (B.P.F. Lelieveldt).

[1] Shared first authorship.

http://dx.doi.org/10.1016/j.ymeth.2014.10.004
1046-2023/

highly-specialized regional markers, poses great challenges for computational approaches. Univariate approaches involving the analysis of the expression profiles of few genes of interest using prior knowledge of their site of action in the brain are not suitable to capture the full complexity of the data. In order to capture the complex patterns of expression of thousands of genes across the entire brain (thousands of samples), multivariate approaches should be employed to accommodate the high-dimensionality of the data. However, visualizing high-dimensional data for intuitive interpretation is challenging.

Several studies have used Principal Component Analysis (PCA) or classical Multidimensional Scaling (cMDS) to reduce the dimensionality of the voxel level genome-wide gene expression data of the mouse brain [7–9]. These low-dimensional maps are then used either to enable visual exploration of the gene expression patterns or as an input to a clustering algorithm where the resulting clusters are compared to the classical neuroanatomy. Classical methods such as PCA and cMDS focus on appropriately modeling large pairwise distances between gene expression profiles [10]. The focus on modeling large pairwise distances comes at the price of substantial errors in modeling small pairwise distances. However, it is exactly this local similarity structure that is essential in clustering and visual exploration: the goal of clustering is to find groups of nearby data points and, similarly, the goal of visual exploration is generally to determine which parts of the data are similar to a reference data point [11]. Therefore, we advocate to employ embedding techniques that focus on preserving local similarity structure, as is done by techniques such as t-Distributed Stochastic Neighbor Embedding (t-SNE) [12]. Since its introduction in 2008, t-SNE has been proven to outperform linear dimensionality reduction methods, but also non-linear embedding methods such as ISOMAP [13], in several research fields including machine-learning benchmark datasets and hyper-spectral remote sensing data [14].

Recently, t-SNE has been employed to analyze high dimensional proteomic and genomic data. Shekhar et al. [15] used t-SNE to differentiate between cellular phenotypes of the immune system based on mass cytometry data. Ji [16] used t-SNE to analyze the relationship between gene expressions and neuroanatomy in the developing mouse brain showing that t-SNE is able to capture the local similarities in the high-dimensional space. Fonville et al. [17] have shown that t-SNE outperforms PCA and self-organizing maps when used for modeling of mass spectrometry imaging data, where each pixel represents a molecular mass spectrum. All the previously mentioned applications demonstrate the high potential of t-SNE in the visual analysis of high-dimensional molecular data.

The goal of this work is to explore the effectiveness and limitations of t-SNE for spatial mapping of gene expression patterns in both the mouse and the human Allen Brain Atlases. By applying Barnes-Hut-SNE (BH-SNE) [22], a recently developed optimization algorithm for t-SNE, we show the consistency of the low dimensional embedding across the 6 human brains as well as between the sagittal and coronal experiments of the mouse brain. In addition, we quantitatively show the superiority of BH-SNE over PCA and cMDS in separating neuroanatomical regions in the low-dimensional 2D embeddings. Finally, we assess the effect of higher-order principal components on the local and global structure of the spatial transcriptome similarity maps.

2. Materials and methods

2.1. Mouse brain gene expression

The Allen Mouse Brain Atlas [6,18] provides genome-wide cellular-resolution *in situ* hybridization (ISH) gene expression data for approximately 20,000 genes of the 8-week old adult C57BL/6 J male mouse brain. For each gene, sagittal ISH sections were sampled at 25 μm intervals across the entire brain and the high-resolution 2D image series from each experiment were reconstructed in 3D and registered to the Nissl stain-based reference atlas (Allen Reference Atlas). The data were then aggregated into isotropic voxels defined by a uniform 200 μm grid in the reference space by averaging the expression levels and densities of all pixels (in the high-resolution ISH sections) within each voxel. The ontology of the reference atlas is used to label individual voxels with their anatomical nomenclature. In addition, coronal sections are available for a set of approximately 4000 genes that showed marked regional expression patterns in the sagittal plane [3]. More information about the ISH sections alignment and registration to the Allen Reference Atlas can be found in [19].

We retrieved all expression energy volumes from [18] using the Allen Brain Atlas application programming interface (API). Expression energy is a measurement combining the expression level (the integrated amount of signal within each voxel) and the expression density (the amount of expressing cells within each voxel) [20].

We focused our analysis on a subset of high confidence genes for which coronal and sagittal experiments are available, as in [8]. For each gene, we computed the Spearman's rank correlation between the corresponding coronal and sagittal experiments and selected genes in the top-three quartiles of correlation (3241 genes). The coronal and sagittal experiments corresponding to those 3241 genes were retained for further analysis (Supplementary Table 1). For genes with more than one sagittal experiment, the maximum correlation value was used. A mask was applied to exclude all non-brain voxels, resulting in a $61{,}164 \times 3241$ (voxels $\times$ genes) matrix for the coronal experiments and a $27{,}365 \times 3241$ matrix for the sagittal experiments.

2.2. Human brain gene expression

The Allen Human Brain Atlas [7,21] includes RNA microarray data collected from the postmortem brains of six donors, with no known neuropsychiatric or neuropathological history; see Table 1 for detailed information about the donors. Magnetic resonance (MR) T1-weighted (T1W), T2-weighted (T2W) and Diffusion Tensor (DT) images were collected in-cranio, prior to dissection for anatomic visualization of each brain.

Approximately 1000 samples were dissected using manual macrodissection for large regions and laser captured microdissection for smaller regions from two donor brains (H0351.2001 and H0351.2002), representing all structures across the whole brain. For the other four donor brains, approximately 500 samples were taken from one hemisphere only. Each sample is associated with a 3D (x, y, z) coordinate on its corresponding donor's MRI volume. Moreover, the MNI coordinates of each sample is reported (registration to the MNI reference space was done using FreeSurfer software). The dataset contained expression profiles of 29,191 genes represented by 58,692 probes, with 93% of known genes represented by at least 2 probes. The data was already normalized across samples and across different brains using the procedure

Table 1
Human donors information.

Donor ID	Number of samples	Sex	Age (years)	Race/ethnicity
H0351.2001	946	Male	24	African American
H0351.2002	893	Male	39	African American
H0351.1009	363	Male	57	Caucasian
H0351.1012	529	Male	31	Caucasian
H0351.1015	470	Female	49	Hispanic
H0351.1016	501	Male	55	Caucasian

explained in [22]. Probes with no Entrez ID or gene symbol were excluded and the expression profiles of all the probes representing one gene were averaged, resulting in 20,737 genes (Supplementary Table 2).

2.3. Cell type markers

Lists of cell-type specific genes were extracted from a previously published work by Cahoy et al. [23] who profiled gene expression patterns in purified populations of neurons, astrocytes, and oligodendrocytes using microarrays. We selected genes enriched by at least 10-fold in one cell type, compared to the two other cell types, out of the 3241 high-confidence genes included in our analysis, resulting in 195 neuron-specific, 60 astrocyte-specific, and 43 oligodendrocyte-specific genes in the mouse data (Supplementary Table 3). Using the same 10-fold enrichment threshold for the human data resulted in 247 neuron-specific, 151 astrocyte-specific, and 92 oligodendrocyte-specific human orthologous genes (Supplementary Table 4).

2.4. Non-linear dimensionality reduction

Three different multivariate data analysis methods were used to visualize the high dimensional expression data, namely: Barnes-Hut-SNE (see Section 3), Principal Component Analysis (PCA) and classical Multi-Dimensional Scaling (cMDS). For the mouse data, the expression profile of each gene, *i.e.* column of the voxel $\times$ gene expression matrix, was Z-score normalized across all voxels. The human brain expression data was already Z-score normalized [22]. PCA was applied to the expression matrices (voxel $\times$ gene for the mouse data and sample $\times$ gene for the human data). The human data was also analyzed with cMDS. The first two components of PCA and the first two dimensions of cMDS were used to visualize the data in each case. The goodness-of-fit criterion for cMDS was the stress, normalized by the sum of squares of the inter-point distances [24]. Distances within cMDS between two samples s_1 and s_2 were computed as genetic distances: $d(s_1,s_2) = (1-\rho(s_1,s_2)^2)^{1/2}$, where $\rho(s_1,s_2)$ denotes the correlation between gene expression levels.

For the non-linear dimensionality reduction using Barnes-Hut-SNE (BH-SNE), we used the full data dimensionality and mapped it to a 2D BH-SNE plot. In addition, we assessed the effect of prior dimensionality reduction using PCA in order to reduce noise in the final maps. The data was first mapped either to the first 2, 3, 5, 10, or 20 components and then embedded into 2D BH-SNE maps.

2.5. Regional gene expression visualization

For both the mouse and the human data, the mapped data points (voxels or samples) were colored in the low-dimensional 2D map according to their associated reference atlas ontology colors, as obtained from the mouse and human atlases [6,7] (Supplementary Tables 5 and 6). This ontology was colorized so that each brain structure has a unique color and anatomically related structures (*e.g.*, substructures of the hypothalamus) are coded with similar colors.

To visually analyze the ability of the different methods (BH-SNE, PCA, and MDS) to segment different regions of the mouse brain, the data points (voxels) were colored by spanning an "L*a*b*" color map [25], that maps the a* and b* colormap axes to the horizontal and vertical axes of the PCA, cMDS and BH-SNE maps. The L*a*b* color space was selected because it spans all perceivable colors and the a* and b* axes span a two-dimensional space with all perceivable colors at a constant perceived "lightness" L* that is perceptually linear. L* was fixed at 50 for all plots providing a good color contrast in the 2D maps.

Using the MNI152 coordinates associated with each of the human brain samples, we mapped each sample back to the Colin27 human brain atlas [26]. Direct visualization of these samples is, however, hampered by the spatial sparsity of the data, *i.e.* there are very few samples per anatomical regions. Therefore, we colored each voxel in the brain where there is no sample available (no gene expression data) according to the closest sampled voxel based on the Euclidean distance between the unsampled voxels and the sampled voxel.

2.6. Evaluation of the mapped data

To evaluate the capability of each of the dimensionality reduction methods to separate different brain regions, we analyzed the separation between different brain-region clusters in the low-dimensional 2D maps produced by these methods. The separation between two brain structures can be characterized by the similarity of the distributions of the data points belonging to each region in the low-dimensional space. In this work we use the Jensen–Shannon divergence [27] to compute this similarity. Briefly, for each brain structure, a 2D histogram is computed by calculating the density of the data points belonging to that structure in the 2D map. This is achieved by overlaying a 40×40 rectangular grid that covers all the samples in the low dimensional space. The divergence between the two histograms (distributions) of two brain structures P and Q is then calculated as: $JSD(P||Q) = \frac{1}{2}KL(P||M) + \frac{1}{2}KL(Q||M)$, where $KL(P||Q) = \sum_{i \neq j} p_{ij} \log \frac{p_{ij}}{q_{ij}}$ is the Kullback–Leibler divergence between P and Q, and $M = \frac{1}{2}(P+Q)$. By using the same grid size (the same number of bins), divergences between brain regions are comparable between the different dimensionality reduction methods regardless of their scaling factor. For the mouse data, divergence scores were calculated between: cortex (Isocortex), olfactory areas (OLF), hippocampal formation (HPF), cortical subplate (CTXsp), striatum (STR), pallidum (PAL), cerebellum (CB), thalamus (TH), hypothalamus (HY), midbrain (MB), pons (P), and medulla (MY). In the human atlas, divergences were calculated between: frontal lobe (FL), parietal lobe (PL), temporal lobe (TL), occipital lobe (OL), hippocampal formation (HiF), striatum (Str), Globus pallidus (Gp), amygdala (Amg), thalamus (TH), hypothalamus (Hy), mesencephalon (MES), Pons, myelencephalon (MY), cerebellum (Cb), and white matter (WM).

3. Theory

3.1. Barnes-Hut-SNE (BH-SNE)

t-Distributed Stochastic Neighbor Embedding (t-SNE) [12] constructs a two-dimensional scatter plot in which each point represents a gene expression profile. In such a t-SNE map, nearby points correspond to similar profiles, whereas distant points correspond to dissimilar profiles. The map is constructed by (1) measuring similarities between gene expression profiles and (2) moving points around in the map in such a way as to minimize some difference measure between similarities of points in the map and the corresponding gene-expression profile similarities.

Mathematically, t-SNE operates by converting the gene expression profiles into a probability distribution over pairs of profiles in such a way, that similar pairs have a high probability of being picked. The distribution is defined as a standard Gaussian kernel (with a particular choice for σ^2) that is normalized to sum to one. Next, t-SNE constructs a map in which each point corresponds to an expression profile by: (1) defining a similar distribution over the pairs of points in the map, and (2) minimizing the divergence

between the two distributions with respect to the coordinates of the points in the map using gradient descent. Mathematically, for a given sample t-SNE defines pairwise similarities between gene expression profiles $\boldsymbol{x} \in \mathbb{R}^D$ (with D the number of genes) as:

$$p_{ij} = \frac{\exp(-||\boldsymbol{x}_i - \boldsymbol{x}_j||^2/2\sigma^2)}{\sum_{k \neq l} \exp(-||\boldsymbol{x}_k - \boldsymbol{x}_l||^2/2\sigma^2)}$$

where p_{ij} is the similarity between the expression profiles $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ in the high dimensional space $\mathbb{R}^D$. Likewise, the similarity between the corresponding low-dimensional models of these genes $\boldsymbol{y} \in \mathbb{R}^d$ (with d the dimensionality of the t-SNE map) is defined as:

$$q_{ij} = \frac{(1 + ||\boldsymbol{y}_i - \boldsymbol{y}_j||^2)^{-1}}{\sum_{k \neq l}(1 + ||\boldsymbol{y}_k - \boldsymbol{y}_l||^2)^{-1}}$$

where q_{ij} is the similarity between the expression profiles $\boldsymbol{y}_i$ and $\boldsymbol{y}_j$ in the low dimensional space $\mathbb{R}^d$. In the definition of map similarity, a heavy-tailed Student-t distribution is used to measure similarity in the map to account for the large difference in volume between the high-dimensional gene expression space and the low-dimensional map. The low-dimensional map $\{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N\}$, with N the number of gene expression profiles, is learned by minimizing the Kullback–Leibler divergence between both distributions:

$$KL(P||Q) = \sum_{i \neq j} p_{ij} \log \frac{p_{ij}}{q_{ij}}$$

The asymmetry of the Kullback–Leibler divergence encourages modeling large P-values (similar expression profile) by large Q-values (nearby points). As a result, in contrast to techniques like PCA, t-SNE focuses on appropriately modeling the local gene expression profile structure in the map.

The gradient that is used for learning a t-SNE map can be interpreted as an N-body system [28]: each point in the map exerts a force onto all other points, and the gradient for each point is the resultant force on that point (*i.e.* the sum of all incoming forces). Specifically, the gradient on point $\boldsymbol{y}_i$ comprises springs between $\boldsymbol{y}_i$ and all other points, where the force in a spring depends on the difference between the corresponding P-values and Q-values. The gradient computes the resultant force on map point $\boldsymbol{y}_i$ in this spring system:

$$\frac{\partial KL}{\partial \boldsymbol{y}_i} = 4 \sum_{j \neq i} (p_{ij} - q_{ij})(1 + ||\boldsymbol{y}_i - \boldsymbol{y}_j||^2)^{-1} (\boldsymbol{y}_i - \boldsymbol{y}_j)$$

The interpretation of t-SNE as an N-body simulation facilitates the use of approximation techniques that were originally developed in astronomy to simulate large galaxies of stars, such as the Barnes-Hut approximation [29] or fast multipole approximations [30]. We focus on the Barnes-Hut approximation [28][2], which exploits the fact that a group of nearby map points exerts very similar forces on another point that is relatively far away. Therefore, the (resultant) force exerted on the latter point can be approximated by the force between the center-of-mass of the group of points and the point under consideration, multiplied by the number of points in the group. In practice, this is implemented by storing all points in a quadtree (for 2D maps) or octree (for 3D maps) and performing a depth-first tree search on this tree, pruning away nodes for which the aforementioned approximation can be used. The resulting algorithm has an average-case complexity of $O(N \log N)$. For more details, we refer to [28].

[2] A variant of t-SNE based on fast multipole methods is presented in [31].

4. Results and discussion

4.1. Genetic similarity within the mouse brain

We used BH-SNE to embed the mouse coronal expression data in a 2D space, see Fig. 1a. In order to reduce the noise in the data, we first reduced the data dimensionality by mapping the data on the first 10 principal components and then used the reduced data as an input to BH-SNE. The BH-SNE mapped data show that anatomical regions are in many cases in disjoint and visually distinct clusters. By comparison, Fig. 1d plots the reduced data when only PCA is being used, showing that then samples of the same anatomical region are close but no clear clustering and regional separation is visible. This difference between BH-SNE and PCA is also reflected in Fig. 1b and e, where the borders between anatomical regions seem sharper with coloring based on BH-SNE. In Fig. 1b for example the hippocampal formation (in red) and cerebral nuclei (in light blue) can be easily distinguished. More strikingly, the transversal views in Fig. 1b shows that within the deeper brain structures there is a clear difference between medial gene expression (green) and posterior gene expression (pink). The first 2 principal components do not pick up on this variation and suggest a strong similarity (Fig. 1e). To quantify these observations, Fig. 1f shows the Jensen–Shannon divergence between classical anatomical regions in both the first two principal components and the two-dimensional BH-SNE. A lower divergence value indicates that the map suggests higher similarity between the corresponding brain regions in the 2D map. In general, BH-SNE yields much higher divergence values between pairs of anatomical regions compared to PCA. Particularly, the cortex (Isocortex) and the hippocampus (HPF) are not separable in the PCA map, that is why they both retain the same color (pink) in Fig. 1e. This high similarity between the cortex and hippocampus is also reflected with a low divergence value in Fig. 1f. On the other hand, BH-SNE can clearly separate the cortex and the hippocampus giving them different colors in Fig. 1b and resulting in a high divergence value in Fig. 1c.

Adding more PCA components in the BH-SNE preprocessing step increases the information available to the BH-SNE algorithm and thereby influences the separation of neuroanatomical regions in the embedded map. To study the effect of PCA preprocessing step on the BH-SNE mapping we varied the number of principal components. Fig. 2 shows the BH-SNE plots and the L*a*b* mappings to the mouse brain (transversal sections) using a range of components to reduce the dimensionality of the data before applying BH-SNE. When using a small number of components (5 or less), insufficient information is retained to achieve good separation between different brain regions. This is visible from the mixing of samples from the same anatomical region in the BH-SNE maps and in the axial brain slices by ragged edges (Fig. 2a and b, *e.g.*, yellow, blue and green clusters). By increasing the number of components to 5 or 10, BH-SNE produces much better results as seen by clear separation between voxels belonging to different brain structures (Fig. 2c and d). When we further increased the number of components (20 or more), the clusters of voxels belonging to one brain structure started to break into smaller sub-clusters. This is particularly clear in Fig. 2f (BH-SNE without prior dimension reduction). These sub-clusters seem to be formed by voxels belonging to different coronal planes, visible when the points in the BH-SNE map are projected back to the mouse brain (Fig. 2f, lower panel). When the BH-SNE map is colored according to the coronal plane from which the data was extracted, the observed sub-clusters could indeed be attributed to the different coronal planes (Fig. 2g). Since the gene expression data was generated using ISH if coronal section of the mouse brain, we can reason that by including more components BH-SNE starts to pick up inter-slice

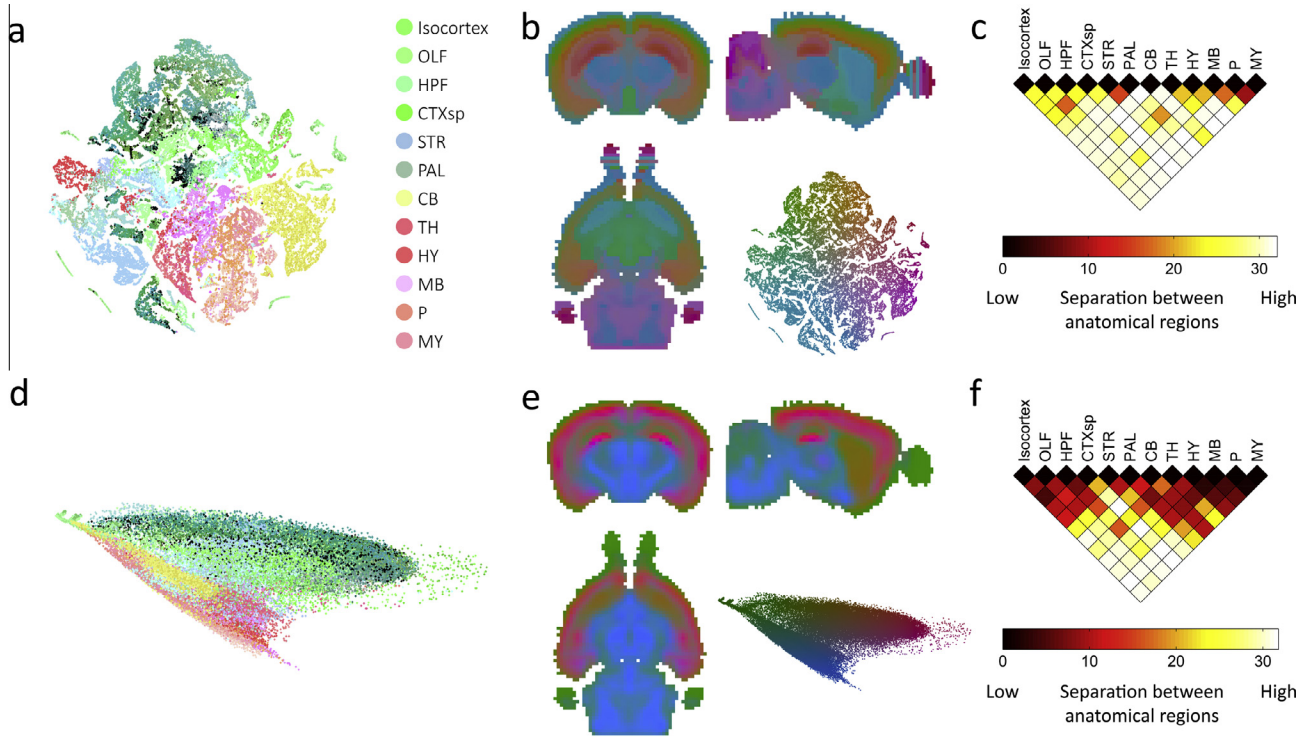

Fig. 1. Coronal mouse brain transcriptome similarities. (a) BH-SNE map of the mouse coronal data initialized with 10 principal components and colored by anatomical region labels from the Allen Reference Atlas: cortex (Isocortex), olfactory areas (OLF), hippocampal formation (HPF), cortical subplate (CTXsp), striatum (STR), palladium (PAL), cerebellum (CB), thalamus (TH), hypothalamus (HY), midbrain (MB), pons (P), and medulla (MY). (b) The mouse coronal data mapped back to the 3D volume of the mouse atlas (3 views) and colored by the L*a*b* colormap of the BH-SNE mapping at a constant L* value. (c) Divergence plot for BH-SNE showing the similarity between pairs of neuroanatomical regions. A higher divergence value (lighter colors) indicates better separation between a pair of neuroanatomical regions in the 2D BH-SNE map. (d) The first two PCA components of the mouse coronal data colored by anatomical region labels from the Allen Reference Atlas. (e) The mouse coronal data mapped back to the 3D volume of the mouse atlas (3 views) and colored by the L*a*b* colormap of the PCA mapping. (f) Divergence plot for PCA showing the similarity between pairs of neuroanatomical regions. A higher divergence value (lighter colors) indicates better separation between a pair of neuroanatomical regions in the 2D PCA map.

differences. Neighboring voxels within the same brain structure and within the same slice have a more similar expression profile than neighboring voxels within the same brain structure but in different coronal slices. This leads to the fragmentation of brain region clusters into smaller, within-slice clusters, and hence the visible color gradient over the coronal brain slices (Fig. 2f).

Ko et al. [9] demonstrated that k-means clustering on cell-type specific genes reveals that neuron-specific genes show the most neuroanatomically similar pattern across the mouse brain. To explore if cell type composition gives rise to the expression differences that separate anatomical regions, we performed BH-SNE on cell type marking genes only. Fig. 3 shows the BH-SNE mappings for three disjoint sets of genes. The neuron-specific genes yield the highest separation between neuroanatomical regions (Fig. 3a). In order to quantify the ability of different gene sets to partition neuroantomy we use the Jensen–Shannon divergence plot. The neuron-specific gene set leads to stronger separation of the cortex (Isocortex), olfactory area (OLF), hippocampus (HPF), and the cortical subplate (CTXsp) as well as between the thalamus (TH) and the hypothalamus (HY), indicated by higher divergence values (lighter color) in Fig. 3d–f. Overall oligodendrocyte-specific genes show the weakest separation between classical anatomical regions, although all gene sets give clear distinctions between cortex and non-cortex areas. This confirms the observations reported in [9] on k-means clustering based on expression of cell-type specific genes.

The BH-SNE analysis was also performed on the sagittal data, which spans a smaller volume of interest than the coronal data (27,365 voxels compared to 61,164). This map (in Fig. 4a and b) is highly comparable to that of the coronal experiments, with especially clear distinctions between cerebral nuclei (medium blue) and cerebral cortex (orange). In this case, selecting fewer principal components for the BH-SNE initialization does not lead to slicing effects as observed in the coronal samples, but it does emphasize within-slice similarities which are visible by a smoother color transitions in Fig. 4c.

4.2. Genome wide gene expression similarity within the human brain

In the human brain atlas platform paper, Hawrylycz et al. [7] used PCA and cMDS mappings to show that the transcriptional relationships between cortical samples mimic the spatial topography of the cortex. To visualize the anatomical organization of the high-dimensional expression data through the entire adult human brain, we mapped the data to a 2D map using BH-SNE without prior dimensionality reduction. By mapping the expression data of each of the six brains separately, we observed that BH-SNE is able to map the samples with clear clustering of samples belonging to the same anatomical region, see Fig. 5. Particularly, the cortex (red, yellow and brown samples) and the cerebellum (light blue) are clearly separated from all other brain regions across the six brains, indicating that both regions have distinct expression profiles from the rest of the brain. Moreover, the thalamus (light green), hippocampus (ochre) and the caudate nucleus and putamen (purple) are consistently close to each other in the low dimensional space, indicating that the expression profiles of these regions are more similar to each other than to other regions of the brain. Remarkably, the caudate nucleus and putamen are clustered, while the third organ of the basal ganglia, the globus pallidus, has a separate cluster. The relationships between samples in the low-dimensional space are very consistent across the six brains, suggesting a global organization of the human brain

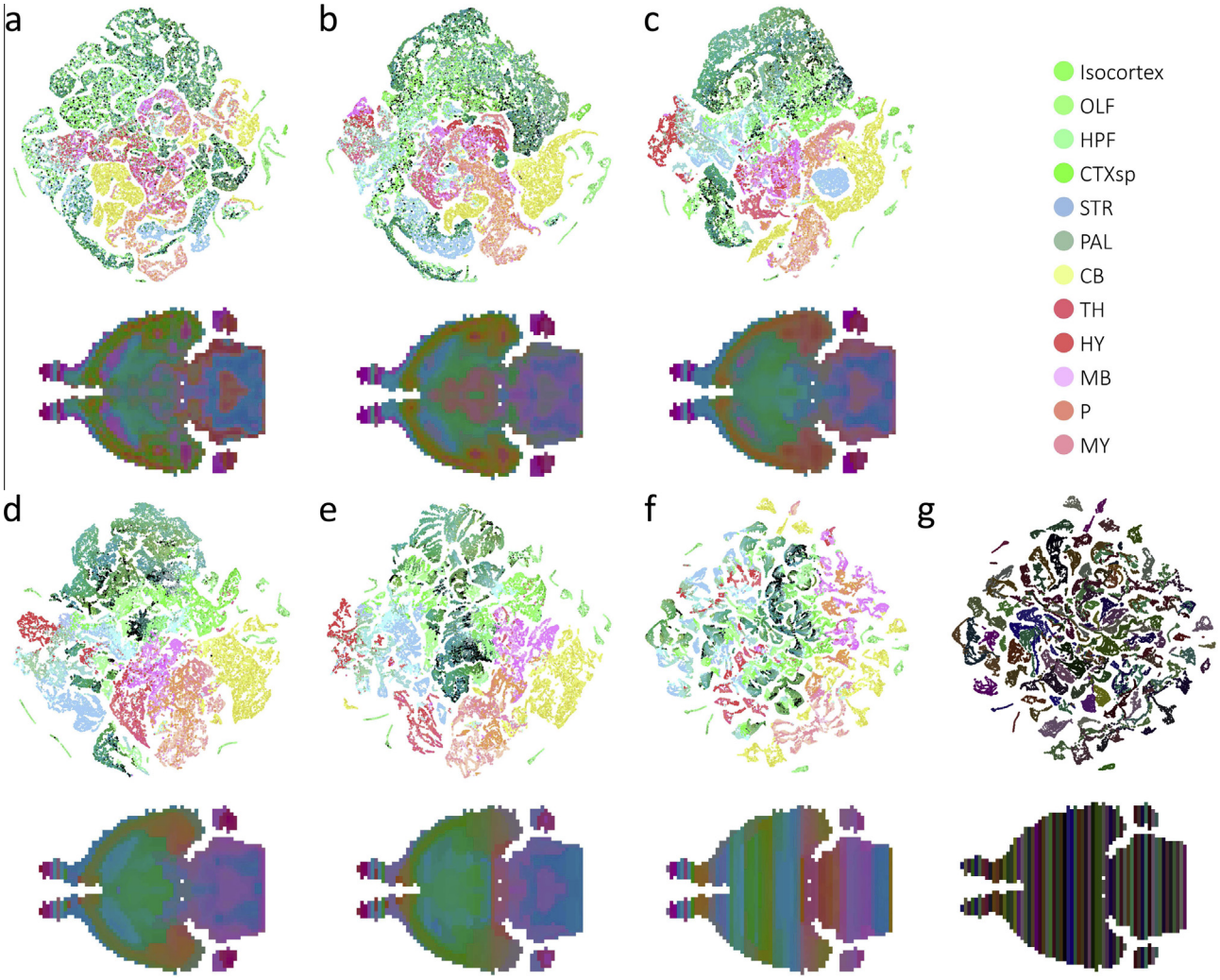

Fig. 2. BH-SNE maps the mouse coronal data using different initializations. BH-SNA mappings of the mouse coronal data using (a) 2, (b) 3, (c) 5, (d) 10 and (e) 20 principal components to reduce the dimensionality of the data before applying BH-SNE. (f) BH-SNE mapping of the mouse coronal data without any prior dimensionality reduction. BH-SNE maps are colored by anatomical region labels from the Allen Reference Atlas. Axial sections of the mouse 3D atlas space are colored with the corresponding L*a*b colors of each voxel in the BH-SNE maps. (g) BH-SNE mapping of the mouse coronal data without any prior dimensionality reduction, colored by the corresponding coronal plane.

transcriptome across individuals. It is worth noting that BH-SNE optimizes the pair-wise distances between pairs of data points, but not the absolute location of each data point in the 2D map. Therefore, when analyzing the maps in Fig. 5, one should consider the relative distances between samples from different neuroanatomical regions rather than the absolute geometric location within the map.

To compare the similarity of the expression profiles of different brain regions to the classical neuroanatomy, we compared the $L^*a^*b^*$ colors of the BH-SNE mapping to the original structural labels on the MNI152 atlas for Donor H0351.2001 and Donor H0351.2002, from whom both hemispheres were sampled in the ABA. In Fig. 6, the differences between cortical, cerebellar, and brain stem samples are clearly visible. However, we could not identify differences between the frontal, medial, and anterior regions of the cortex in the t-SNE map. On the other hand, regions surrounding the ventricles clearly differ from the adjacent brain regions and there are no clear differences within the brainstem. These maps reveal the global symmetry between hemispheres in regional gene expression that was also reported in [7].

To gain insight into the effect of cell-type specific genes on the mapping, we inspected embeddings of human brain samples when creating the maps using only the expression profiles of cell-type specific genes. Similar to our observation from the mouse data, neuron-specific genes encode better separability between anatomical regions (Fig. 7a) with a BH-SNE map very similar to the BH-SNE map obtained using the entire set of genes. Furthermore, astrocyte-specific genes resulted in a more separable map compared to oligodendrocyte-specific genes (Fig. 7b and c). In all three mappings, there is a strong overlap between the frontal lobe (FL), temporal lobe (TL), occipital lobe (OL), and parietal lobe (PL) (Fig. 7d–f), which was also visible in the brain maps in Fig. 6. It is also worth noting that the overall differences between the cell-type specific maps are much smaller in the human data that the corresponding maps in the mouse data (Fig. 3).

We then pooled all the samples from the six brains and mapped them to the low-dimensional 2D space using BH-SNE (with prior dimensionality reduction to 10 principal components), PCA, and cMDS. The BH-SNE map of the concatenated data resembled those of the individual donors to a large extent, with samples from the cortex and the cerebellum clearly separated from samples in the rest of the brain, see Fig. 8. Again, BH-SNE (Fig. 8a) retains a better separation between samples belonging to different anatomical regions as compared to PCA (Fig. 8b) and cMDS (Fig. 8c). For PCA, the overlap between the cerebellum (light blue) and the cortex (red, yellow and brown) can be resolved when the 3rd component

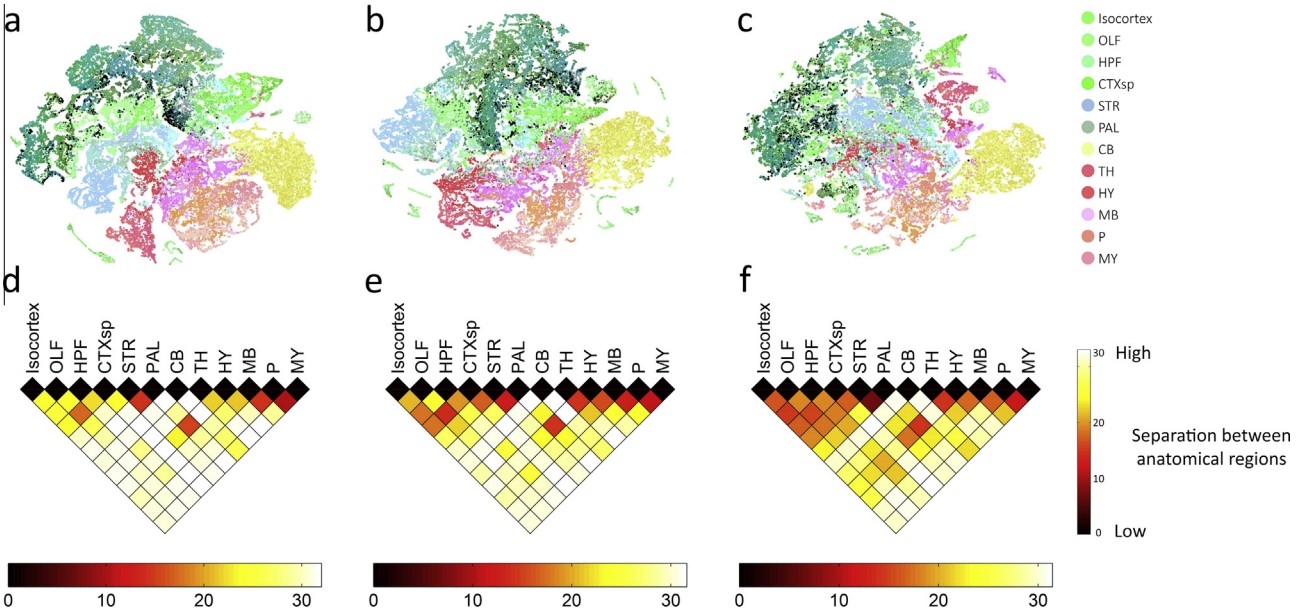

Fig. 3. BH-SNE embeddings of cell-type specific genes in the mouse coronal data. (a) neuron-specific, (b) astrocyte-specific, and (c) oligodendrocyte-specific gene sets based BH-SNE embeddings, colored by anatomical region labels from the Allen Reference Atlas. (d–f) Divergence plots of the BH-SNE embeddings using (d) neuron-specific, (e) astrocyte-specific, and (f) oligodendrocyte-specific gene sets. A higher divergence value (lighter colors) indicates better separation between a pair of neuroanatomical regions in the 2D BH-SNE map.

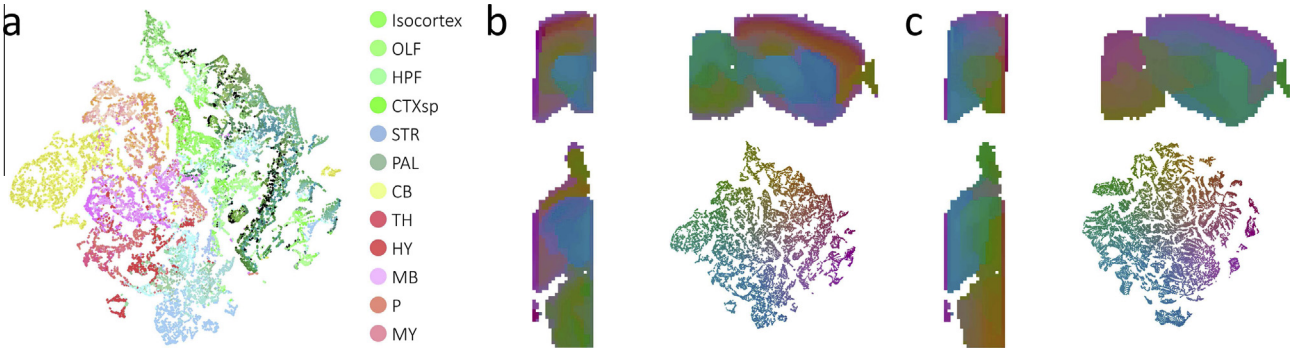

Fig. 4. BH-SNE maps of the mouse sagittal data. (a) BH-SNE map of the mouse sagittal data using 10 principal components for the initial dimensionality reduction and colored by anatomical region labels from the Allen Reference Atlas. (b) Mapping of the BH-SNE embedding back to the 3D volume of the mouse atlas (3 views), colored by the L*a*b* colormap of the BH-SNE embedding. (c) Mapping of the BH-SNE embedding using the high dimensional sagittal data without prior dimensionality reduction.

is taken into account. Within the cerebellum, none of the three methods could separate frontal (FL), temporal (TL), occipital (OL) and parietal lobes (PL), consistent with our findings from individual brains (Fig. 7). This further supports the superiority of BH-SNE to retain variations in higher components in the 2D space. Separation between donors becomes apparent in BH-SNE (Fig. 8d), which is much less apparent in PCA and cMDS. In order to quantify the separation between different anatomical structures in the low dimensional space, we computed the Jensen–Shannon divergences between the regions; see Fig. 8g–i. The divergence plots show clearly that the BH-SNE map has higher divergence values, hence retaining a better separation between all neuroanatomical regions followed by PCA and subsequently cMDS. The sensitivity of BH-SNE is also demonstrated by its ability to distinguish samples from different donors. Fig. 8a and d show different clusters per brain region and per donor. In the PCA and cMDS mappings, samples from different donors are fully mixed.

In Fig. 2f, we have shown that by retaining more PCA components, one can separate the mouse expression data based on inter-slice differences. To analyze the effect of retaining more PCA components prior to the BH-SNE embedding in the human brain, we gradually increased the number of principal components used to initialize the BH-SNE mapping. Fig. 9 (top row) shows that by increasing the number of principal components, *i.e.* increasing the data dimensionality, before applying BH-SNE, samples belonging to the same anatomical structure, but to different donors, start to deviate from each other. When colored according to the source brain, Fig. 9 (bottom row), deviations in the BH-SNE maps appear to reflect differences between brains only when more components are used in the prior dimensionality reduction. At 5 components (Fig. 9c) we start to observe a separation of the samples from H0351.2001 brain regions (red) and H0351.2002 brain regions (yellow) from the other samples, especially in the cortical and cerebellar regions. The other four brains (H0351.2009, H0351.2012, H0351.2015, and H0351.2016) become clearly separated when much higher components are included (20 components, Fig. 9e). However, BH-SNE is still able to maintain the separation between different anatomical structures even when higher components are added, but at the costs that each brain region in each donor then forms its own cluster. The clustering in anatomy related

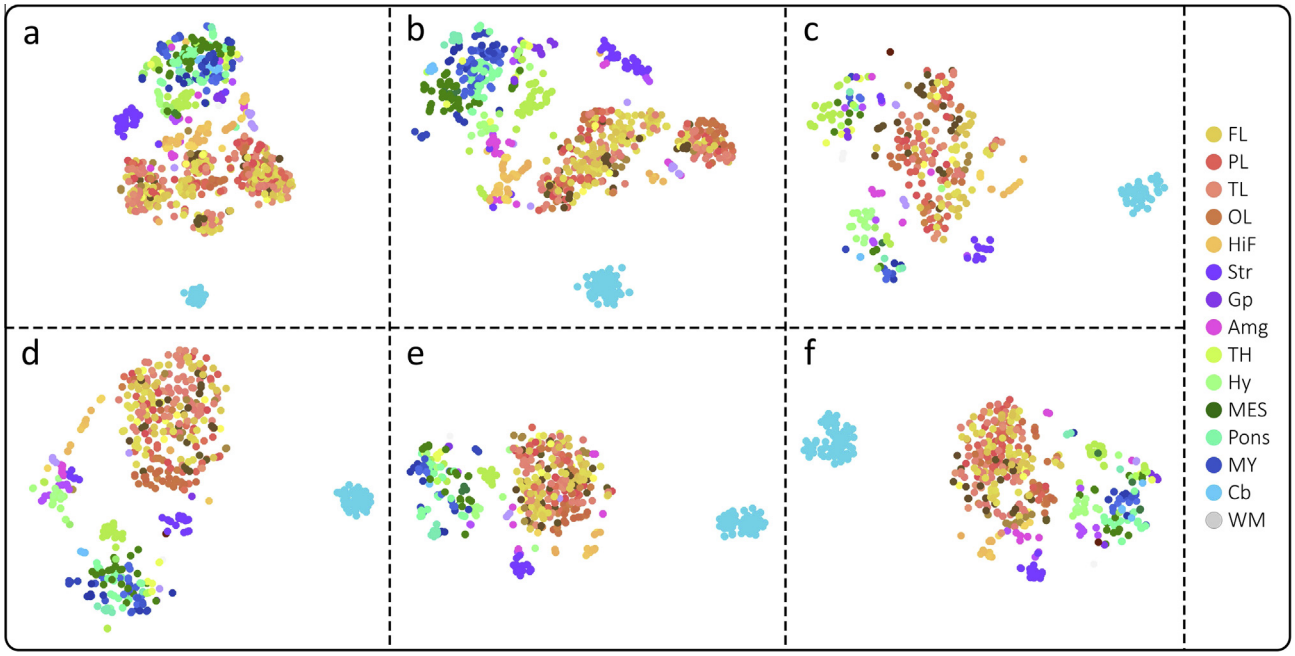

Fig. 5. Genome-wide gene expression similarity within six human brains. BH-SNE maps of the expression data for each of the six human brain donors individually without prior dimensionality reduction by PCA. (a) Donor H0351.2001. (b) Donor H0351.2002. (c) Donor H0351.2009. (d) Donor H0351.2012. (e) Donor H0351.2015. (f) Donor H0351.2016. BH-SNE maps are colored by anatomical region labels from the Allen Reference Atlas: frontal lobe (FL), parietal lobe (PL), temporal lobe (TL), occipital lobe (OL), hippocampal formation (HiF), striatum (Str), Globus pallidus (Gp), amygdala (Amg), thalamus (TH), hypothalamus (Hy), mesencephalon (MES), Pons, myelencephalon (MY), cerebellum (Cb), and white matter (WM). Note the higher density in (a) and (b) is due to the larger number of samples.

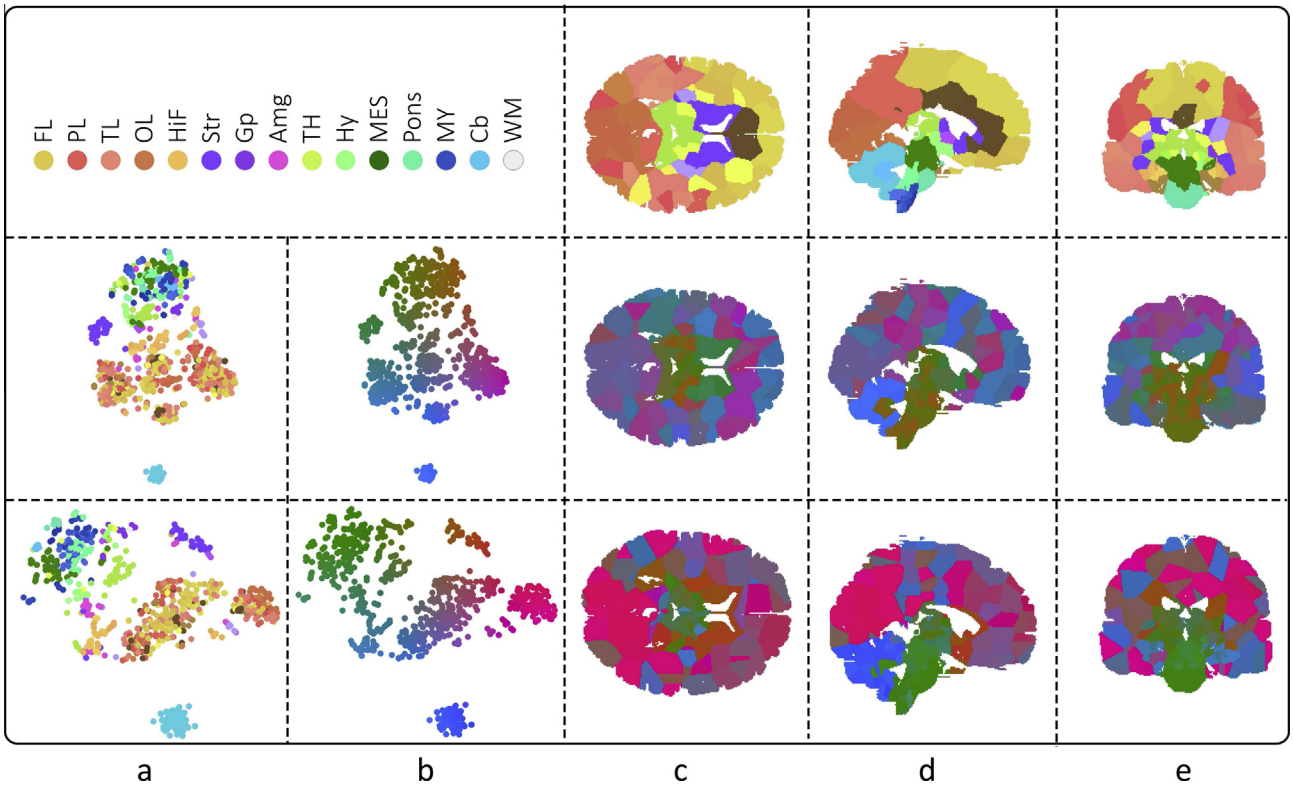

Fig. 6. Anatomical view on the genome-wide gene expression similarity within two human brains. BH-SNE maps of the expression data without prior dimensionality reduction by PCA, (a) colored by anatomical region labels from the Allen Reference Atlas (see legend) and (b) colored by the L*a*b* colormap. (c) Transverse, (d) sagittal, and (e) coronal views of the brain colored according to the L*a*b* colors of the samples in the BH-SNE maps. Top row: Anatomical labels from the ABA projected on the full brain. Middle row: The brain of H0351.2001. Bottom row: The brain of H0351.2002. Note the clear separation of the cerebellum and brain stem in (b) as well as that the global preservation of the symmetry in regional gene expression between hemispheres in (c and e).

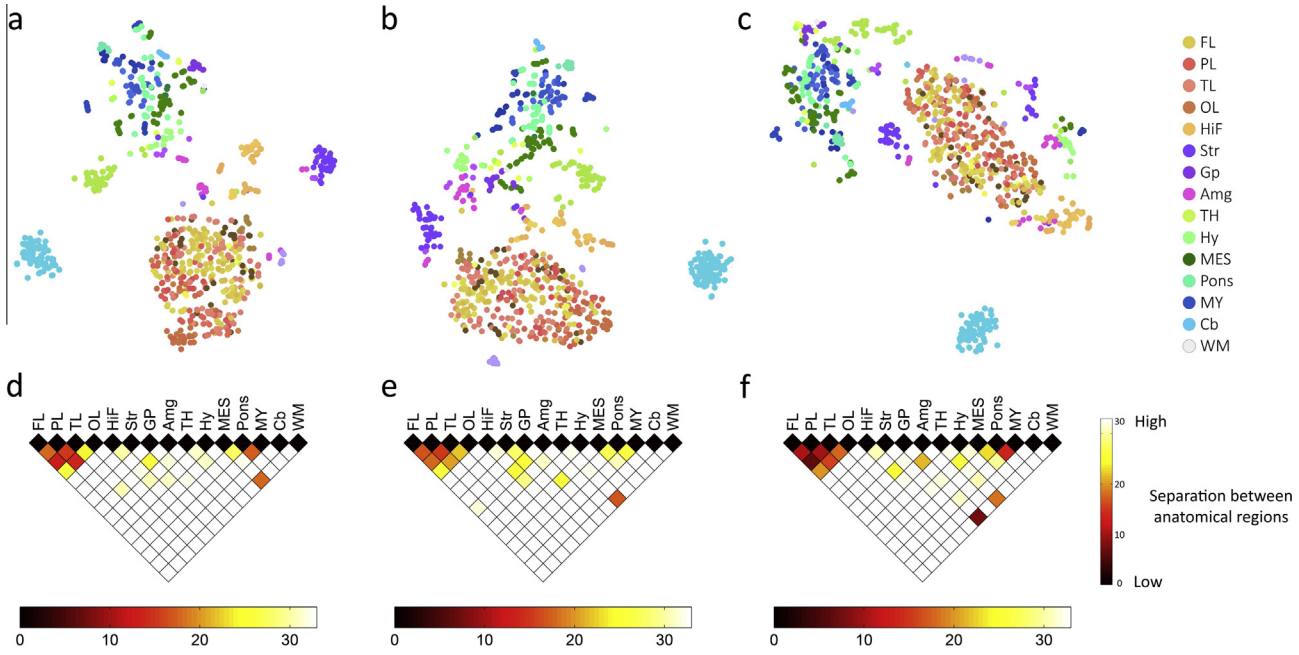

Fig. 7. BH-SNE mappings of the human brain based on cell-type specific genes. BH-SNE embeddings of samples from donor H0351.2002 using (a) neuron-specific, (b) astrocyte-specific, and (c) oligodendrocyte-specific gene sets, colored by anatomical region labels from the Allen Reference Atlas. (d–f) Divergence plots of the BH-SNE embeddings using (d) neuron-specific, (e) astrocyte-specific, and (f) oligodendrocyte-specific gene sets. A higher divergence value (lighter colors) indicates better separation between a pair of neuroanatomical regions in the 2D BH-SNE map.

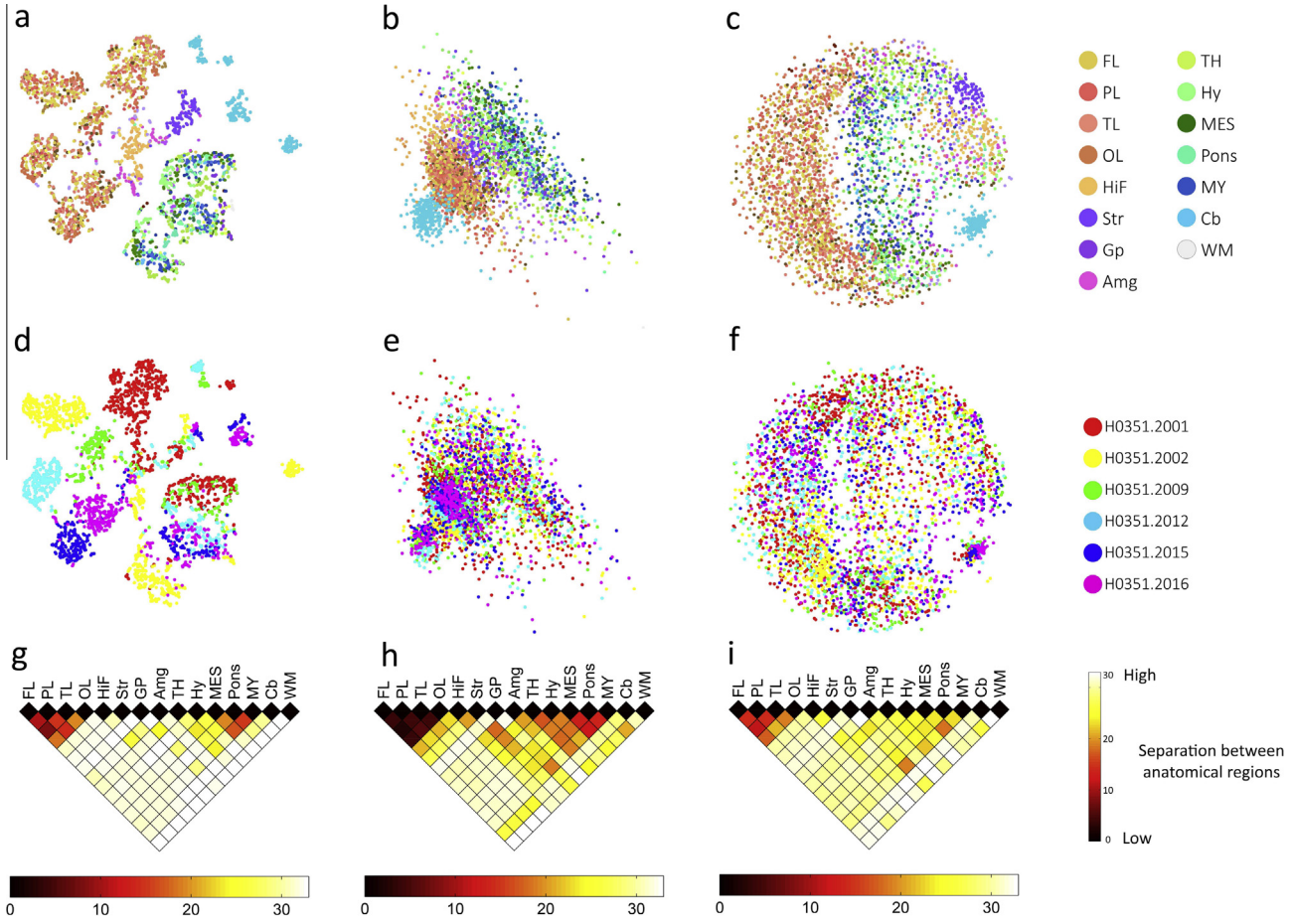

Fig. 8. Linear and non-linear embeddings of the human brain transcriptome. (a) BH-SNE using 10 principal components for prior dimensionality reduction, (b) PCA, and (c) cMDS embeddings of the aggregated gene expression data of the six human brains. Maps a–c are colored by anatomical region labels from the Allen Reference Atlas, maps d–f are colored by donor. (g–i) Jensen–Shannon divergence plots between neuroanatomical regions for BH-SNE, PCA and cMDS, respectively.

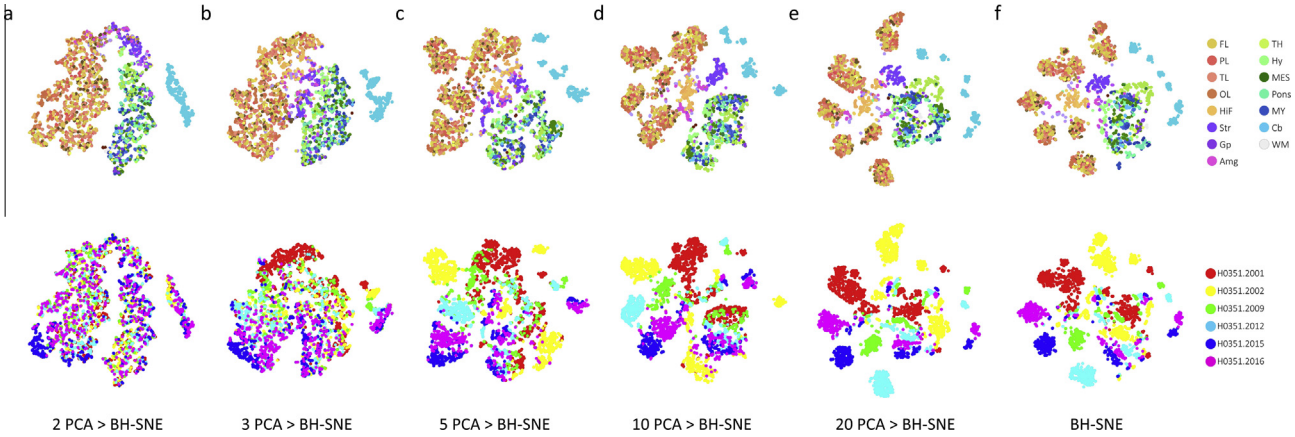

Fig. 9. BH-SNE maps of the human data when using different PCA initializations. BH-SNE embeddings of the aggregated data from six human brains when using (a) 2, (b) 3, (c) 5, (d) 10 and (e) 20 principal components before applying BH-SNE. (f) BH-SNE mapping of the human data without any prior dimensionality reduction. (Top row) BH-SNE maps colored by anatomical region labels from the Allen Reference Atlas. (Bottom row) BH-SNE maps colored by donor brain.

clusters for a low number of retained components shows that the variations in normalized gene expression levels between brain regions are dominant over the variations between donors.

5. Conclusions

We have explored the effectiveness of using t-Distributed Stochastic Neighbor Embedding (t-SNE) to assess the spatial organization of genome-wide expression data across the mammalian brain. We have used Barnes-Hut-SNE (BH-SNE), a recently developed, computationally efficient t-SNE optimization algorithm, to map the large volumes of data in the mouse and human Allen Brain Atlases. Our results show that the mapped gene-expression data is highly consistent between the coronal and sagittal mouse atlases as well as between the six human brain datasets, with the cortex and cerebellum always being the most distinct from other brain regions. Additionally, the BH-SNE maps of the human brain show clear expression symmetry between hemispheres. The separation of neuroanatomical regions in the BH-SNE embedding is better than the separation in the PCA and MDS embeddings, further supporting the need for non-linear embedding methods to capture the complex organization of the Allen Brain Atlas data. We have employed the Jensen–Shannon divergence to quantify the ability of different gene sets and different embedding methods to map brain samples in 2D while preserving the known neuroanatomy. Additionally, we studied the effect of keeping more PCA components prior to the BH-SNE mapping. Due to its emphasis on local structures, BH-SNE is sensitive to having more PCA components as input, even when they may encode for non-anatomical information such as inter-slice differences in the coronal mouse data and different donor brains in the human data. These results suggest that to map high-dimensional spatial transcriptome data to a two dimensional space, a combination of a linear PCA mapping followed by a non-linear BH-SNE mapping gives the best tradeoff between preserving local and global structure in one 2D map.

Acknowledgments

This research has received partial funding from the Dutch Technology Foundation STW, as part of the STW project 12721: "Genes in Space" under the IMAGENE perspective program, and from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 604102. The authors gratefully acknowledge the Allen Institute for Brain Science.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ymeth.2014.10.004.

References

[1] L.W. Swanson, Brain Architecture: Understanding the Basic Plan, second ed., Oxford University Press, New York, 2011.
[2] M. Hawrylycz, L. Ng, D. Page, J. Morris, C. Lau, S. Faber, et al., Neural Netw. 24 (2011) 933–942, http://dx.doi.org/10.1016/j.neunet.2011.06.012.
[3] L. Ng, A. Bernard, C. Lau, C.C. Overly, H.-W. Dong, C. Kuan, et al., Nat. Neurosci. 12 (2009) 356–362, http://dx.doi.org/10.1038/nn.2281.
[4] L. Richardson, S. Venkataraman, P. Stevenson, Y. Yang, J. Moss, L. Graham, et al., Nucleic Acids Res. 42 (2014), http://dx.doi.org/10.1093/nar/gkt1155.
[5] The Gene Expression Nervous System Atlas (GENSAT) Project, NINDS Contracts N01NS02331 & HHSN271200723701C to The Rockefeller University, New York, NY, 2003.
[6] E.S. Lein, M.J. Hawrylycz, N. Ao, M. Ayres, A. Bensinger, A. Bernard, et al., Nature 445 (2007) 168–176, http://dx.doi.org/10.1038/nature05453.
[7] M.J. Hawrylycz, E.S. Lein, A.L. Guillozet-Bongaarts, E.H. Shen, L. Ng, J.A. Miller, et al., Nature 489 (2012) 391–399, http://dx.doi.org/10.1038/nature11405.
[8] J.W. Bohland, H. Bokil, S.D. Pathak, C.-K. Lee, L. Ng, C. Lau, et al., Methods 50 (2010) 105–112, http://dx.doi.org/10.1016/j.ymeth.2009.09.001.
[9] Y. Ko, S.A. Ament, J.A. Eddy, J. Caballero, J.C. Earls, L. Hood, et al., Proc. Natl. Acad. Sci. U.S.A. 110 (2013) 3095–3100, http://dx.doi.org/10.1073/pnas.1222897110.
[10] C.K.I. Williams, Mach. Learn. 46 (2002) 11–19, http://dx.doi.org/10.1023/A:1012485807823.
[11] J. Venna, J. Peltonen, K. Nybo, H. Aidos, S. Kaski, J. Mach. Learn. Res. 11 (2010) 451–490.
[12] L. Van Der Maaten, G. Hinton, J. Mach. Learn. Res. 9 (2008) 2579–2605, http://dx.doi.org/10.1007/s10479-011-0841-3.
[13] J.B. Tenenbaum, V. de Silva, J.C. Langford, Science 290 (2000) 2319–2323, http://dx.doi.org/10.1126/science.290.5500.2319.
[14] D. Lunga, S. Prasad, M.M. Crawford, O. Ersoy, IEEE Signal Process. Mag. 31 (2014) 55–66, http://dx.doi.org/10.1109/MSP.2013.2279894.
[15] K. Shekhar, P. Brodin, M.M. Davis, A.K. Chakraborty, Proc. Natl. Acad. Sci. U.S.A. 111 (2014) 202–207, http://dx.doi.org/10.1073/pnas.1321405111.
[16] S. Ji, BMC Bioinformatics 14 (2013) 222, http://dx.doi.org/10.1186/1471-2105-14-222.
[17] J.M. Fonville, C.L. Carter, L. Pizarro, R.T. Steven, A.D. Palmer, R.L. Griffiths, et al., Anal. Chem. 85 (2013) 1415–1423, http://dx.doi.org/10.1021/ac302330a.
[18] Allen Institute for Brain Science. Allen Mouse Brain Atlas [Internet], (2014). Available from: <http://mouse.brain-map.org/>.
[19] L.Ng,S.D.Pathak,C.Kuan,C.Lau,H.Dong,A.Sodt,etal.,IEEE/ACMTrans.Comput.Biol. Bioinform. 4 (2007) 382–393, http://dx.doi.org/10.1109/tcbb.2007.1035.
[20] C. Lau, L. Ng, C. Thompson, S. Pathak, L. Kuan, A. Jones, et al., BMC Bioinformatics 9 (2008) 153, http://dx.doi.org/10.1186/1471-2105-9-153.
[21] Allen Institute for Brain Science. Allen Human Brain Atlas [Internet], (2014). Available from: <http://human.brain-map.org/>.
[22] Allen Human Brain Atlas Technical White Paper: Microarray Data Normalization, 2013.
[23] J. Cahoy, B. Emery, A. Kaushal, L. Foo, J. Zamanian, K. Christopherson, et al., J. Neurosci. 28 (2008) 264–278, http://dx.doi.org/10.1523/JNEUROSCI.4178-07.2008.
[24] T.F. Cox, M.A.A. Cox, Multidimensional Scaling, Chapman & Hall, 1994.

[25] G. Sharma, Digital Color Imaging Handbook, 1.7.2 ed., CRC Press, 2003.
[26] C.J. Holmes, R. Hoge, L. Collins, R. Woods, A.W. Toga, A.C. Evans, J. Comput. Assist. Tomogr. 22 (2) (1998) 324–333, http://dx.doi.org/10.1097/00004728-199803000-00032.
[27] A.R. Webb, Statistical Pattern Recognition, second ed., John Wiley and Sons Ltd., 2002.
[28] L.J.P. van der Maaten, J. Mach. Learn. Res. (2014).
[29] J. Barnes, P. Hut, Nature 324 (1986) 446–449, http://dx.doi.org/10.1038/324446a0.
[30] L. Greengard, V. Rokhlin, J. Comput. Phys. 73 (1987) 325–348, http://dx.doi.org/10.1016/0021-9991(87)90140-9.
[31] M. Vladymyrov, M. Carreira-Perpinan, Linear-time training of nonlinear low-dimensional embeddings, in: 17th Int. Conf. Artif. Intell. Stat. (AISTATS 2014), 2014, pp. 968–977.