Title

Chromatic fusion: Generative multimodal neuroimaging data fusion provides multi-informed insights into schizophrenia.

Journal

Human Brain Mapping, 44(17)

Authors

Plis, Sergey
Calhoun, Vince
Geenjaar, Eloy
et al.

Publication Date

2023-12-01

DOI

10.1002/hbm.26479

Peer reviewed

Received: 9 March 2023 | Revised: 4 August 2023 | Accepted: 24 August 2023

DOI: 10.1002/hbm.26479

RESEARCH ARTICLE

WILEY

Chromatic fusion: Generative multimodal neuroimaging data fusion provides multi-informed insights into schizophrenia

Eloy P.T. Geenjaar[1,2] | Noah L. Lewis[2,3] | Alex Fedorov[1,2] | Lei Wu[2] | Judith M. Ford[4,5] | Adrian Preda[6] | Sergey M. Plis[2,7] | Vince D. Calhoun[1,2,3,7,8]

[1]School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, Georgia, USA

[2]Tri-Institutional Center for Translational Research in Neuroimaging and Data Science (TReNDS), Georgia State, Georgia Tech, Emory, Atlanta, Georgia, USA

[3]School of Computational Science and Engineering, Georgia Institute of Technology, Atlanta, Georgia, USA

[4]San Francisco Veterans Affairs Medical Center, San Francisco, California, USA

[5]Department of Psychiatry and Behavioral Sciences, University of California San Francisco, San Francisco, California, USA

[6]Department of Psychiatry and Human Behavior, University of California Irvine, Irvine, California, USA

[7]Department of Computer Science, Georgia State University, Atlanta, Georgia, USA

[8]Department of Psychology, Georgia State University, Atlanta, Georgia, USA

Correspondence
Eloy P.T. Geenjaar, School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, USA.
Email: egeenjaar@gatech.edu

Funding information
National Science Foundation, Grant/Award Number: 2112455; National Institutes of Health, Grant/Award Number: R01EB006841; Georgia Tech/Emory NIH/NIBIB Training Program, Grant/Award Number: T32EB025816

Abstract

This work proposes a novel generative multimodal approach to jointly analyze multimodal data while linking the multimodal information to colors. We apply our proposed framework, which disentangles multimodal data into private and shared sets of features from pairs of structural (sMRI), functional (sFNC and ICA), and diffusion MRI data (FA maps). With our approach, we find that heterogeneity in schizophrenia is potentially a function of modality pairs. Results show (1) schizophrenia is highly multimodal and includes changes in specific networks, (2) non-linear relationships with schizophrenia are observed when interpolating among shared latent dimensions, and (3) we observe a decrease in the modularity of functional connectivity and decreased visual-sensorimotor connectivity for schizophrenia patients for the FA-sFNC and sMRI-sFNC modality pairs, respectively. Additionally, our results generally indicate decreased fractional corpus callosum anisotropy, and decreased spatial ICA map and voxel-based morphometry strength in the superior frontal lobe as found in the FA-sFNC, sMRI-FA, and sMRI-ICA modality pair clusters. In sum, we introduce a powerful new multimodal neuroimaging framework designed to provide a rich and intuitive understanding of the data which we hope challenges the reader to think differently about how modalities interact.

KEYWORDS
deep learning, *multimodal fusion*, static functional connectivity, structural/functional/diffusion magnetic resonance imaging, variational autoencoders, visualization

1 | INTRODUCTION

The acquisition of neuroimaging data often consists of various modalities, such as structural, functional, and diffusion magnetic resonance imaging data. Access to a wide range of complementary modalities is necessary to deeply understand the relationship between (functional/structural) brain patterns and demographic/neuropsychiatric variables (Calhoun & Sui, 2016). For example, unimodal studies may show that

brain activity and structure are both linked to the same neuropsychiatric variable. This, however, does not allow us to draw the conclusion that both modalities (linearly or nonlinearly) covary together. In fact, combining modalities may have an even stronger link to the neuropsychiatric variable under study (Liu et al., 2015). Understanding the covariation pattern of two modalities and how that covariation relates to neuropsychiatric variables also allows us to dig deeper into the latent mechanism that underlies both the divergence in behavioral and neuroimaging measures. When covariation in divergent brain activity and structure are linked to the same brain region, the effects may be localized. However, when covariation indicates different brain regions for each modality, there may be a more complex mechanism underlying their effect on the neuropsychiatric variable. Importantly, jointly estimating the multimodal relationships can allow the modalities to inform one another, yielding information that is richer than a unimodal analysis. A rich understanding of these patterns will lead to complementary quantitative visualizations or measures that have the potential to help clinicians make decisions about diagnosis and treatment plans. Further, access to quantitative visualizations will ultimately be important to help practitioners facilitate the delivery of personalized medicine (Bhugra et al., 2017).

Multimodal neuroimaging is thus a critical aspect of studying the brain and aims to incorporate each modality as a piece of the puzzle to obtain a complete picture of the brain. Just as with pieces of a landscape puzzle, modalities have exclusive or private features (e.g., the picture of a specific tree or house printed on the puzzle piece) and shared features (e.g., the neighborhood that emerges only when pieces are combined). By thinking of features from a variety of neuroimaging modality pairs as private and shared, they become more interpretable and also allow us to learn more structured features from the modalities. We exploit the more conceptual interpretability of private and shared features in our framework, since we now have three dimensions along which subjects can differ, two sets of private features (one for each modality), and one set of shared features. We use the likeliness of the three sets of features in the dataset to compare subjects to each other. This is visualized in our framework as a color for each subject, based on the three feature sets, which correspond to red, green, and blue. Thus allowing us to map each subject's multimodal feature sets to a color with a higher intensity for unlikely feature sets, and a low intensity for likely feature sets compared to the rest of the data. However, what features are shared and what features are private between pairs of neuroimaging modalities is not known apriori. We therefore utilize recent deep learning work that has started to develop neural networks to extract private and shared features from pairs of modalities (Lee & Pavlovic, 2021; Shi et al., 2019). Specifically, our multimodal neuroimaging model leverages and extends the disentangled multimodal variational autoencoder (DMVAE) (Lee & Pavlovic, 2021) as a building block. The goal of our work is to learn low-dimensional features from multimodal datasets in such a way that we can understand how the interactions among the modalities in each dataset relate to schizophrenia. We chose to use an approach that models two modalities as having private and shared features because it induces a structure in our low-dimensional representation that both constrains the solution space of our algorithm and is interpretable. Furthermore, we can map the features in the representation back into the modality's original space, which makes these representations more interpretable. We expect that explicitly modeling shared features between modalities with a non-linear method leads to interesting features potentially related to schizophrenia.

By reducing the number of features and increasing the interpretability of sets of features, we can actually analyze these features more deeply because they are constrained in dimensionality. This allows us to consider subjects on our new constrained spectrum, which we do by assigning colors to each subject based on the three sets of features as mentioned previously. Moving toward multi-dimensional (continuous) measures to understand psychiatric disorders is important because binary labels can be misleading. For example, schizophrenia often co-occurs with other mental disorders (Upthegrove et al., 2017). Additionally, there is significant intra-diagnostic heterogeneity for schizophrenia, and the lines between other severe mental disorders remain blurred. Thus, representing neuroimaging features on a constrained multi-dimensional spectrum is important to understand individual brain differences, and predict risk (Sui et al., 2020). Integrating multiple modalities and viewing subjects on a spectrum is consistent with the trans-diagnostic NIMH research domain criteria (RDoC) initiative (Sanislow et al., 2019).

In this work, we propose a new framework to analyze spectrum psychiatric disorders through a powerful multimodal representation learning framework. Specifically, we disentangle private and shared features from a pair of modalities and characterize schizophrenia patients from the perspective of these distinct features. Given their variational inference framework, variational autoencoders enforce a certain prior during training. In our case, we use a zero-mean normal distribution, which allows us to naturally interpret encoded modalities further from this prior than others to be more irregular in the dataset. Based on the three distinct types of features; the private features from the first and second modalities, and their shared features, we can assign a color to a subject based on how irregular each of these types of features are for that specific subject. To make this more robust, instead of assigning the color to each subject, we first find clusters in the low-dimensional space and assign a color to each cluster. We then use the colors of the clusters to assign colors to the full spectrum of the space. Given that there are three sets of features, we use a red-green-blue (RGB) color model to assign a color to each cluster, where red and blue represent irregularity in the sets of private features from the first and second modalities, respectively. Green represents irregularity in their shared set of features. We call this framework of creating a low-dimensional multimodal color spectrum along which subjects vary: chromatic fusion. Assigning colors allows us to draw inferences about the dataset quickly, and increases visual interpretability.

In our analyses, we use all potential pairs of four modalities; spatial ICA maps, static functional connectivity (sFNC), fractional anisotropy maps (FA), and voxel-based morphometry (sMRI). Although sFNC and spatial ICA maps are derived from the same modality, namely functional magnetic resonance imaging (fMRI), we refer to them as different modalities in this manuscript for brevity. Our goal is to both introduce a new way of thinking about combining modalities,

namely in terms of colors related to private and shared features, and visualizing how those combinations of modalities uniquely highlight subgroups and thus the heterogeneity of psychiatric spectrum disorders, such as schizophrenia. Specifically, for each modality pair, we highlight the upper quartile (7 out of 27) of naturally arising schizophrenia-enriched clusters; clusters with a large number of schizophrenia subjects, whether those clusters capture the same subjects across modality pairs, and what patterns in the data these clusters represent. Additionally, we look at representative shared features our model finds and how they relate to schizophrenia to highlight the importance of explicitly disentangling shared features from private features. We hypothesize that different modality pairs lead to distinct patterns related to schizophrenia. We aim to find these patterns without supervised labels and compare them across modality pairs. In fact, we show that these naturally arising clusters are a function of which modalities are paired. Our results imply that distinct combinations of modalities are able to highlight different schizophrenia subgroups in our dataset. This also indicates how heterogeneity in schizophrenia is a function of the interactions between modalities. This means that different subgroups are distinct from the rest of the sample based on which modalities are combined. Additionally, we show that some shared features are significantly related to schizophrenia. These shared features naturally arise from our model while training without any supervision and we include analyses on the robustness of these shared features across training folds. Generally, we find robust and correlated shared features for each modality pair, and highlight two to visualize unique and multimodal patterns significantly related to schizophrenia.

1.1 | Previous work

In multimodal machine learning one of the main challenges is to learn good representations (Baltrušaitis et al., 2018). Good representations are essentially accurate and interpretable summaries of a data sample at hand. An example of an impactful summary of neuroimaging data are independent component analysis (ICA) spatial maps or functional connectivity (Garrity et al., 2007; Hutchison et al., 2013). Our method addresses this challenge and performs symmetric multimodal fusion. As opposed to asymmetric fusion, where one modality constrains another, we analyze two modalities concurrently in a symmetric fusion model. Furthermore, we look at how the sub-elements of the modalities are fused after training, and summarize (possibly nonlinear) covariations in the data, instead of trying to find mechanistic relationships between modalities. There are also important distinctions within the field of multimodal data fusion itself (Calhoun & Sui, 2016). Mainly, there are model-driven and data-driven approaches. Model-driven approaches have the advantage of specifying apriori hypotheses that can be tested, and that those hypotheses can be directional. However, if any important hypotheses are missed, these methods can make incorrect inferences. Data-driven approaches on the other hand do not require the specification of a priori hypotheses. Our model falls under data-driven multimodal fusion approaches. Lastly, there is a distinction between blind and semi-blind data-driven fusion approaches (Calhoun & Sui, 2016). Semi-blind approaches, in contrast to blind approaches, use some a priori knowledge to constrain the solution space of the model, such as regularization. As discussed in the previous paragraph, by encouraging our model to find private and shared features for two modalities, we constrain the solution space of the model. Thus, our approach falls under the umbrella of unsupervised semi-blind multimodal fusion approaches. Examples of other semi-blind multimodal fusion approaches are joint ICA (Calhoun et al., 2009), which assumes a shared loading parameter, multiset canonical correlation analysis with reference + joint ICA (mCCAR + jICA) (Qi et al., 2017), which allows a behavioral reference to constrain the solution, independent vector analysis (IVA) (Kim et al., 2006; Lee et al., 2008; Ma et al., 2014), which assumes an independent/dependence structure to extract linked sources, and parallel ICA, which optimizes jointly for independence within a modality, and linear covariation between a subset of the sources. These multimodal fusion approaches are complex and thus indicate there is a potential benefit in leveraging the flexibility of deep learning models in the context of multimodal fusion. Further, we want to move beyond the linear mixing assumption by utilizing a non-linear autoencoder-based network, and explicitly find private and shared features among pairs of modalities.

Two recent studies (Hu et al., 2020; Peide et al., 2022), similar to our approach, model multimodal data in terms of its private and shared features. The authors in the former perform rigorous experiments to find shared and common features using advanced coupled matrix tensor factorization (ACMTF) to predict whether a task is being performed in time windows of simultaneous electroencephalogram-functional magnetic resonance imaging (EEG-fMRI) data. The work extends previous tensor decomposition approaches that have been successful at finding shared features between modalities and fusing them (Silva et al., 2020; Sui et al., 2013). The latter of the two previously mentioned studies uses an autoencoder-based model to disentangle the private and shared features but trains it using a supervised signal to predict infant age. These methods differ from our work both in their goal and approaches. For example, although (Hu et al., 2020) uses a variational autoencoder as well, they train the model adversarially with supervised signals. Our work is completely unsupervised yet also results in common interpretable patterns in the dataset. In addition, no work has to our knowledge exploited these multimodal relationships to capture subgroups and visualized the continuum of multimodal combinations as in our chromatic fusion approach.

Other ways of modeling multimodal neuroimaging data have been proposed as well, two recent approaches involve graph neural networks (Ma et al., 2017; Zhang et al., 2020) to learn multimodal representations from functional magnetic resonance imaging (fMRI) and diffusion MRI (dMRI) data. Alternatively, previous studies have modeled the interaction between fMRI data and structural MRI (sMRI) data by aligning sMRI components with dynamic functional connectivity (Plis et al., 2018). Multimodal learning with sMRI and fMRI data has also been explored using self-supervised learning (Fedorov, Sylvain et al., 2021; Fedorov, Geenjaar et al., 2021), by minimizing the divergence of representations from two different modalities. Our work extends these studies by allowing new points to be visualized in terms of their irregularity along the private and shared dimensions of

the multimodal data and visualized to reveal novel multimodal insights. For example, our results unify co-variations of fractional anisotropy and voxel-based morphometry with increased visual–visual and reduced visual-sensorimotor functional connecitivity, respectively. Furthermore, our approach enables a new way of thinking about modalities in terms of a chromatic framework, which we show can provide additional insights and increase data transparency.

2 | MATERIALS AND METHODS

2.1 | Problem setting

Our goals are to (1) build a generative model that allows us to find private and shared latent features for each modality, and (2) provide a visual representation of the results to facilitate the discovery of additional insights into the relationship between modalities and neuropsychiatric spectrum disorders. To do this, we use the identified sets of features to estimate colors in the latent space and create a chromatically fused color space to describe the dataset. The type of dataset we study in this work $X = \{x^i\}_{i=1}^N$ consists of N subjects, with M modalities per subject $x^i = \left\{m_j^i\right\}_{j=1}^M$. In this work, we consider the case for $M = 2$. One widespread generative model to tackle this problem is the variational autoencoder (VAE) (Kingma & Welling, 2014), which maximizes the log-likelihood of reconstructions in the dataset.

We learn the private and shared sets of features using a modified variational autoencoder, as explained in this section. The data in this problem is assumed to be sampled from an underlying, lower-dimensional, distribution $p(\mathbf{z})$. We assume that the observed data for the two modalities is sampled from the following conditional distributions, one for each modality $p_\theta^1(\mathbf{m}_1^i|\mathbf{z}_1^i)$ and $p_\theta^2(\mathbf{m}_2^i|\mathbf{z}_2^i)$. Where $\mathbf{m}_1^i$ and $\mathbf{m}_2^i$ are the two observed modality samples for subject i. Note that the underlying distribution and these conditional distributions are unknown, but we want to estimate them using the observed data. The estimation of the marginal distributions for the two modalities is intractable because of their form; $\mathrm{p}(\mathbf{m}_1) = \int p(\mathbf{z}_1) p_\theta^1(\mathbf{m}_1|\mathbf{z}_1)\, \mathrm{d}\mathbf{z}_1)$ and $\mathrm{p}(\mathbf{m}_2) = \int p(\mathbf{z}_2) p_\theta^2(\mathbf{m}_2|\mathbf{z}_2)\, \mathrm{d}\mathbf{z}_2)$. The intractability of these marginal distributions also leads to the intractability of the following true posterior densities: $p_\theta(\mathbf{z}_1|\mathbf{m}_1) = \frac{p_\theta^1(\mathbf{m}_1|\mathbf{z}_1)p(\mathbf{z}_1)}{p(\mathbf{m}_1)}$ and $p_\theta(\mathbf{z}_2|\mathbf{m}_2) = \frac{p_\theta^2(\mathbf{m}_2|\mathbf{z}_2)p(\mathbf{z}_2)}{p(\mathbf{m}_2)}$. So instead of estimating these distributions by optimizing over the true posterior densities, the VAE approximates the posterior density for both modalities using an encoder for each $q_\phi^j(\mathbf{z}_j|\mathbf{m}_j)$. This encoder parameterizes a simpler distribution, in our case an axis-aligned Gaussian distribution, than the true posterior. The approximate posterior is variationally optimized to be close to this prior, with zero-mean and diagonal unit covariance. By reparametrizing the conditional distributions, we obtain a variational lower bound on the marginal likelihood of each data point $\mathbf{m}_1^i, \mathbf{m}_2^i \in \mathbf{x}^i$. In the case of VAEs this lower bound is called the evidence lower bound (ELBO) and a more in-depth derivation can be found in Kingma and Welling (2014). In our case, we can write the following equation for the ELBO:

$$\begin{aligned}\mathscr{L}(\theta,\phi;\mathbf{m}_1^i,\mathbf{m}_2^i) = &-D_{KL}\left(q_\phi^1(\mathbf{z}_1|\mathbf{m}_1^i)\,\middle\|\,p(\mathbf{z}_1)\right) + \mathbb{E}_{q_\phi(\mathbf{z}_1|\mathbf{m}_1^i)}\left[\log p^1(\mathbf{m}_1^i|\mathbf{z}_1)\right] \\ &-D_{KL}\left(q_\phi^2(\mathbf{z}_2|\mathbf{m}_2^i)\,\middle\|\,p(\mathbf{z}_2)\right) + \mathbb{E}_{q_\phi(\mathbf{z}_2|\mathbf{m}_2^i)}\left[\log p^2(\mathbf{m}_2^i|\mathbf{z}_2)\right]\end{aligned} \tag{1}$$

This objective function, however, does not account for the interactions between modalities. It optimizes the ELBO for the two modalities separately with two separate encoder-decoder models for each modality. To model the shared features between the two modalities, we model those interactions using a disentangled multimodal VAE (DMVAE) (Lee & Pavlovic, 2021). A visual depiction of the DMVAE is shown in Figure 1b. One of the main conceptual ideas behind the DMVAE (Lee & Pavlovic, 2021) and other multimodal models that have recently gained traction (Shi et al., 2019) is the separation of the latent space into private and shared features for each modality (Baltrušaitis et al., 2018).

Conceptually, some of the features that are captured by different modalities are mutually exclusive, while other features are shared across modalities. The mutually exclusive features a modality captures are generally referred to as its "private" features. For example, T1-weighted structural MRI's (sMRI) can measure cortical thickness in gray matter as its private features, whereas dMRI can act as an index of axonal organization and coherence in white matter (Seitz et al., 2018). These modalities, however, share features regarding white matter. We want to ensure we explicitly model modalities this way to disentangle both private and shared features. Throughout the text, we will refer to the private features of subject i's first modality as $\mathbf{pr}_1^i$, its shared features as $\mathbf{sh}^i$, and the private features of its second modality as $\mathbf{pr}_2^i$. The encoders will still parameterize Gaussian distributions, but their dimensions will be split into private and shared features, see Figure 1b. Thus, we obtain two different shared features, one for each modality, namely $\mathbf{sh}_1^i$ and $\mathbf{sh}_2^i$. To mix these two shared feature sets during training, the authors propose to use a product of experts (PoE) (Hinton, 1999). This allows us to obtain a closed form solution for the variance $\sigma_{\mathbf{sh}^i}$ and mean $\mu_{\mathbf{sh}^i}$ of the combined shared distribution: $\frac{1}{\sigma_{\mathbf{sh}^i}} = \frac{1}{\sigma_{\mathbf{sh}_1^i}} + \frac{1}{\sigma_{\mathbf{sh}_2^i}}$ and $\mu_{\mathbf{sh}^i} = \sigma_{\mathbf{sh}^i}\left(\frac{1}{\sigma_{\mathbf{sh}_1^i}}\mu_{\mathbf{sh}_1^i} + \frac{1}{\sigma_{\mathbf{sh}_2^i}}\mu_{\mathbf{sh}_2^i}\right)$. The private and shared features can be expressed in terms of the encoder $q(\cdot)$ as follows: $q^1\left(\mathbf{pr}_1^i, \mathbf{sh}_1^i|\mathbf{m}_1^i\right)$, $q^2\left(\mathbf{pr}_2^i, \mathbf{sh}_2^i|\mathbf{m}_2^i\right)$, and $p\left(\mathbf{sh}^i|\mathbf{sh}_1^i, \mathbf{sh}_2^i\right)$. The new objective function then consists of 6 reconstruction and 5 KL-divergence terms. $\mathbf{m}_1^i$ $\mathbf{pr}_1^i$

$$\begin{aligned}&\mathscr{L}_{DMVAE}(\theta,\phi;\mathbf{m}_1^i,\mathbf{m}_2^i) \\ &= -D_{KL}\left(q^1(\mathbf{pr}_1^i|\mathbf{m}_1^i)\,\middle\|\,p(\mathbf{z})\right) + \lambda_1\mathbb{E}_{q^1(\mathbf{pr}_1^i,\mathbf{sh}_1^i|\mathbf{m}_1)}\left[\log p^1\left(\mathbf{m}_1^i|\mathbf{pr}_1^i,\mathbf{sh}_1^i\right)\right] \\ &\quad -D_{KL}\left(q^1\left(\mathbf{sh}_1^i|\mathbf{m}_1^i\right)\,\middle\|\,p(\mathbf{z})\right) + \lambda_1\mathbb{E}_{q^1(\mathbf{pr}_1^i|\mathbf{m}_1^i),p(\mathbf{sh}^i|\mathbf{sh}_1^i,\mathbf{sh}_2^i)}\left[\log p^1\left(\mathbf{m}_1^i|\mathbf{pr}_1^i,\mathbf{sh}^i\right)\right] \\ &\quad -D_{KL}\left(q^2(\mathbf{pr}_2^i|\mathbf{m}_2^i)\,\middle\|\,p(\mathbf{z})\right) + \lambda_2\mathbb{E}_{q^2(\mathbf{pr}_2^i,\mathbf{sh}_2^i|\mathbf{m}_2^i)}\left[\log p^2\left(\mathbf{m}_2^i|\mathbf{pr}_2^i,\mathbf{sh}_2^i\right)\right] \\ &\quad -D_{KL}\left(q^2\left(\mathbf{sh}_2^i|\mathbf{m}_2^i\right)\,\middle\|\,p(\mathbf{z})\right) + \lambda_2\mathbb{E}_{q^2(\mathbf{pr}_2^i|\mathbf{m}_2^i),p(\mathbf{sh}^i|\mathbf{sh}_1^i,\mathbf{sh}_2^i)}\left[\log p^2\left(\mathbf{m}_2^i|\mathbf{pr}_2^i,\mathbf{sh}^i\right)\right] \\ &\quad -D_{KL}\left(p\left(\mathbf{sh}^i|\mathbf{sh}_1^i,\mathbf{sh}_2^i\right)\,\middle\|\,p(\mathbf{z})\right) + \lambda_1\mathbb{E}_{q^1(\mathbf{pr}_1^i|\mathbf{m}_1^i),q^2(\mathbf{sh}_2^i|\mathbf{m}_2^i))}\left[\log p^1\left(\mathbf{m}_1^i|\mathbf{pr}_1^i,\mathbf{sh}_2^i\right)\right] \\ &\quad +\lambda_2\mathbb{E}_{q^2(\mathbf{pr}_2^i|\mathbf{m}_2^i),q^1(\mathbf{sh}_1^i|\mathbf{m}_1^i))}\left[\log p^2\left(\mathbf{m}_2^i|\mathbf{pr}_2^i,\mathbf{sh}_1^i\right)\right]\end{aligned} \tag{2}$$

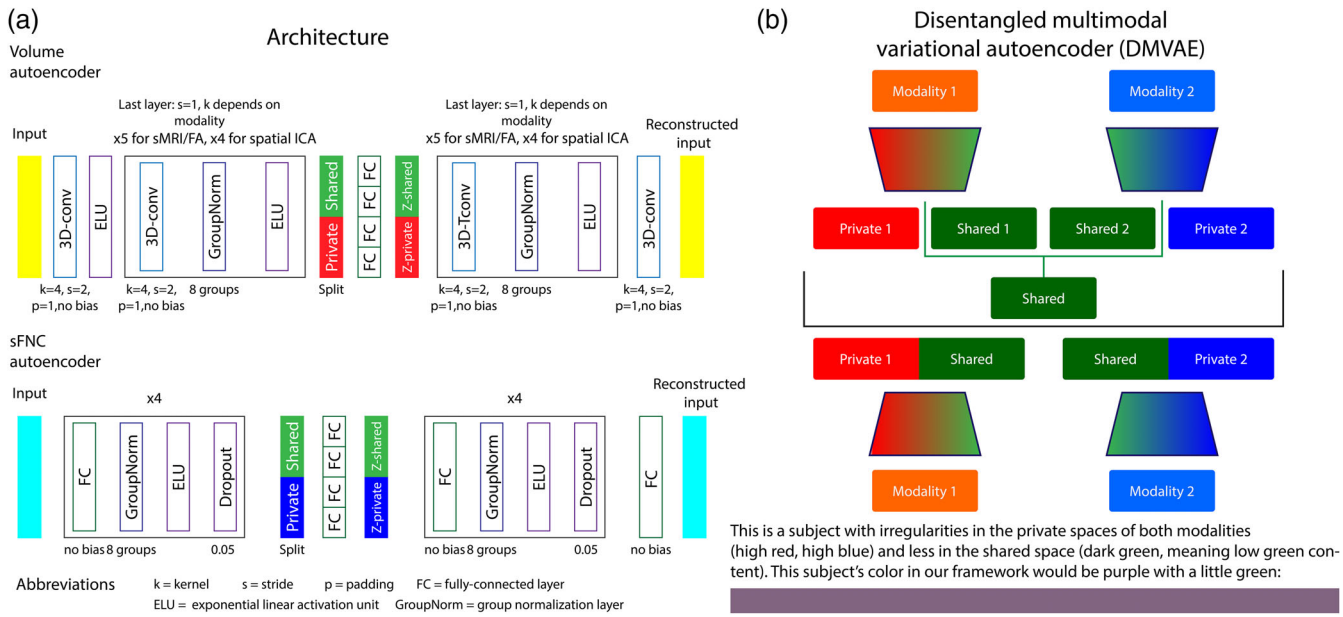

FIGURE 1 The neural network architecture used for each of the modalities. The left part of this diagram shows the architecture of the encoders and decoders used in this work, and the right part shows the high-level structure of the DMVAE and a visual example of how it extracts the colors of a subject's modalities into its base colors.

where $p(\mathbf{z})$ is the diagonal Gaussian prior, λ_1 and λ_2 are weighting factors for the reconstruction of each modality, and the last two reconstruction factors are cross-generation factors that use a sample from the prior of their own private features and a sample from the shared features of the other modality for the cross-reconstruction. Once the model has been trained using this objective function, the shared features of the first modality $\mathbf{sh}_1$ and the second modality $\mathbf{sh}_2$ should have converged toward joint shared features **sh**. Thus, we are left with three separate feature sets; the private features of the first modality $\mathbf{pr}_1$, the private features of the second modality $\mathbf{pr}_2$, and the shared features **sh**. Each subject can be represented as a point in each of these feature sets, this is pictorially represented in Figure 2a. Note that each point, in this case, refers to a normal distribution parameterized by the respective encoder of each modality. The exact implementation of the encoder and decoder for each modality is shown in Figure 1a.

2.2 | Multi-dimensional clustering

After training the model with the objective function in Equation 2, the three feature sets ($\mathbf{pr}_1^i$, $\mathbf{sh}^i$, and $\mathbf{pr}_2^i$) are concatenated together to form one large embedding vector:

$$\text{embedding}^i = \left[\mathbf{pr}_1^i\ \mathbf{sh}^i\ \mathbf{pr}_2^i\right]$$

Each subject is represented by its embedding vector, which we use to cluster subjects and assign colors to the clusters. Since each subject is represented by a multivariate normal distribution, we take its most likely value, the mean, to cluster the subjects. The clusters are then found using K-Means++ clustering in the three feature sets These three feature sets together span a multi-dimensional latent space. A visualization of the clustering is shown in Figure 2b. To understand the relationship among feature sets and subjects we determine the color of each cluster based on their L2-norm. The first private feature set $\mathbf{pr}_1$ is red, the shared feature set **sh** green, and the second private feature set $\mathbf{pr}_2$ is blue. The value for red, blue, and green is determined by the cluster's L2-norm for each feature set, divided by the maximum L2-norm in that feature set. The norm captures how uncommon a feature is along that specific dimension. We can interpret the norm this way because the neural network is trained by minimizing the KL-divergence between each training subject and the prior during training (see Equation 2). Therefore, high KL-divergence is penalized and the model will try to incur KL-divergence penalties as infrequently as possible. A mean further away from 0 (the mean of the prior), will thus incur a penalty during training. The model will try to ensure a high KL-divergence along a certain latent dimension in the feature sets is thus as uncommon as possible. It only uses a high KL-divergence if it aids in the reconstruction error (the other term in the objective function). If not, the magnitude of the KL-divergence would have been optimized to be lower, meaning the subject's mean deviates less from zero. After assigning a color to each cluster, we call the clusters together with their assigned colors metachromatic patterns (MCPs). The "redness" of a given MCP can be interpreted as patterns that have uncommon private features for the first modality. This same interpretation can be applied to the "greenness" and "blueness" of each MCP. After the decoding process, we can define the full latent space in terms of chromatically fused colors, where subjects are samples from this chromatic space. The colors of each individual subject are determined with a probabilistic assignment

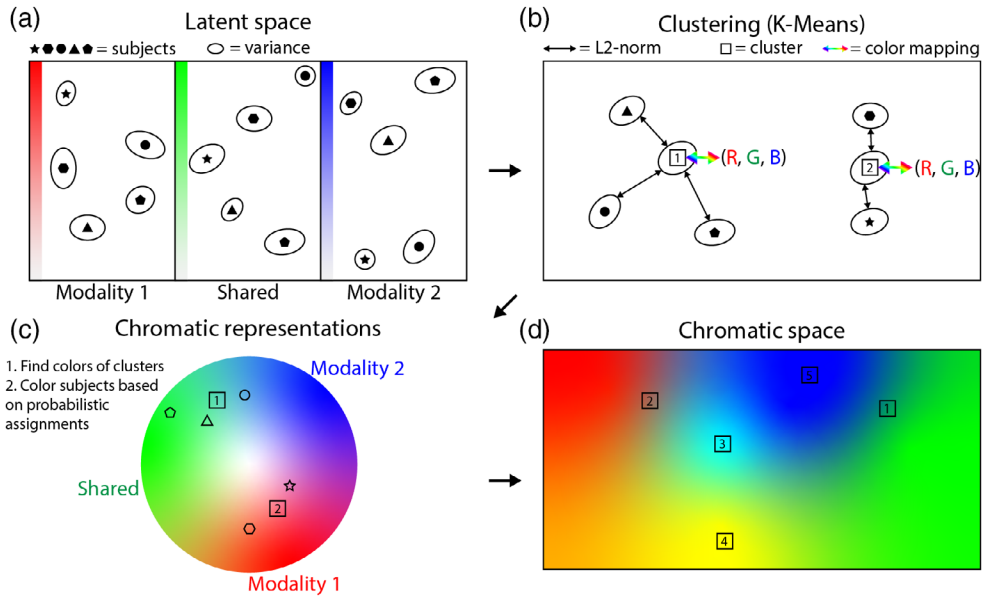

FIGURE 2 The chromatic fusion framework. This figure shows the steps in the complete chromatic fusion algorithm. The top left shows the latent space that is obtained by training a DMVAE. The top right diagram shows the K-means clustering algorithm that is applied to the means of the distributions obtained by the encoder of the DMVAE. Left bottom: the representation of subjects as colors using each meta-chromatic pattern. Right bottom: the final chromatically fused latent space.

to each cluster. This probabilistic assignment is one over the L2 distance to the fourth power between the subject and each cluster, normalized to sum to one. These probabilities are then multiplied by the colors assigned to each MCP. This is pictorially shown in Figure 2c,d. We select the number of MCPs for each modality pair based on the elbow criterion, and show this selection process in Appendix B.

2.3 | Robustness of the multidimensional MCPs

To assess the robustness of the clusters, the dataset is split into 10 stratified folds. Each fold is used as a test set once, and the remaining 9 folds are used to create a training and validation set. The validation set is a random 10% stratified subset of those remaining 9 folds. Validation set subjects are not used to train on, but the objective function (Equation 2) is evaluated on the validation set to ensure the model is not overfitting on the training set. The model with the lowest loss on the validation set is used to cluster the whole dataset. Thus, each cluster occurs 10 times, with different training, validation, and test sets.

We then decode all subjects in our dataset to identify schizophrenia-enriched MCPs in the latent space. In essence, we are trying to establish the robustness of certain subjects being grouped together with respect to changes in the training distribution. To match MCPs across training folds, we take the MCP in the first fold, fold 0. This fold is also used for further visualizations in subsequent sections (see Section 3). The subjects that are assigned to each MCP in fold 0 are then compared to the subjects assigned to each MCP obtained from training and then clustering on the other 9 folds. The overlap between MCPs in fold 0 and the other folds can be expressed as the percentage of subjects that are assigned to both MCPs. We use the overlap as the weight in a linear assignment problem and solve the assignment of MCPs in fold k, $k \neq 0$ to MCPs in fold 0 using the Hungarian algorithm (Kuhn, 1955; Kuhn, 1956). The average percentage of overlap across all the MCPs that were assigned to each MCP in fold 0 indicates the robustness of that MCP with respect to shifts in the training distribution.

The MCPs across the folds are then used to assess specific brain signatures for schizophrenia subjects. By calculating the average percentage of schizophrenia subjects assigned to an MCP across folds, we can understand potential relationships between chromatic colors and subpopulations in the dataset. Given that the model is equipped with a decoder, we can decode schizophrenia-enriched chromatic clusters in the latent space to brain space, and compare them in brain space. The specific hue of the color, as mentioned previously, is a measure of irregularity in terms of the private feature sets and the modalities' shared feature set. The chromatic colors and interpolations between them are thus a visual and informative measure of specific subpopulations in the dataset.

2.4 | Understanding heterogeneity in schizophrenia-enriched MCPs

The MCPs we identify capture distinct relationships among each of the modality pairs, but as we hypothesize (Section 1), we expect different modality pairs to partially capture distinct schizophrenia patients. Thus, after clustering the multi-dimensional latent space for each of the modality pairs, assigning meta-chromatic colors to each of the clusters, and calculating the robustness and percentage of schizophrenia subjects in each MCP, we perform a cross-modality pair analysis. For this analysis, we select schizophrenia-enriched MCPs, at least one for each modality pair, and calculate the percentage of overlap between schizophrenia patients in different clusters, across folds. Once we obtain the percentage of overlap among schizophrenia-enriched MCPs across modality pairs, we visualize the average percentage of unique subjects in an MCP.

2.5 | The importance of the shared features

One important aspect of the model we use is the fact it can capture non-linear co-variations between the two modalities in a modality pair. To qualify the assumption that capture shared features between modalities is essential to understanding complex mental disorders, we visualize shared latent dimensions for two different modality pairs. Additionally, to quantify the importance of the shared features we analyze their stability and correlation to schizophrenia subjects for each of the latent dimensions in a modality pair's shared features. The results and in-depth analysis are provided in Appendix C. We select an interesting latent dimension from two different modality pairs and show the interpolation from the schizophrenia-enriched part of the latent dimension to the control-enriched part of the latent dimension. Given that the shared features span multiple dimensions, we encode the whole dataset into those dimensions and then calculate the correlation between that dimension and schizophrenia. Since there are no guarantees that a model finds the same latent dimension across each training fold, we only consider latent dimensions that have an average correlation of over 0.7 across folds, see Appendix C.

2.6 | Cross-reconstruction

Another way to evaluate the shared features is to quantify how well modalities can be cross-reconstructed from other modalities using the shared features. We can use the shared features of the modality that is present plus the prior of the missing modality to create a cross-reconstruction. The modality that is present for a subject, say modality 1, first encodes the data to create a representation that is split into its private features $\mathbf{pr}_1$, and shared features $\mathbf{sh}_1$. The latter is optimized during training to be similar to the shared features of modality two $\mathbf{sh}_2$. The private features of the second modality are then sampled from the prior $\mathbf{p}(\mathbf{m}_2)$. To reconstruct the missing modality, the shared features of modality one $\mathbf{sh}_1$ are concatenated with the private features sampled from the prior $\mathbf{p}(\mathbf{m}_2)$ and passed to the decoder. The decoder then reconstructs the sample in the modality's original space. Note that the prior in this case is the one we trained the DMVAE with, which is a zero-mean unit Gaussian distribution.

A complete overview of all the steps we perform in our proposed framework is shown in Figure 3.

2.7 | Data

2.7.1 | Acquisition and demographics

The main dataset used in this work is the function bioinformatic research network (fBIRN) phase III data, a schizophrenia dataset (Keator et al., 2016), and the demographics of individuals with each modality pair are described in Table 1. The schizophrenia patients and controls were matched based on age, gender, handedness, and race distributions. Some subjects do not pass the quality assessment of either of the modalities, which means there may be fewer subjects for certain modality pairs. We take the largest number of subjects for each modality pair, such that we maximize the size of the dataset. Interesting in this context is that our method can also generate missing modalities, providing a way to alleviate this issue in the future.

The dataset itself consists of scans collected at seven consortium sites (University of Minnesota, University of Iowa, University of New Mexico, University of North Carolina, University of California Los Angeles, University of California Irvine, and University of California

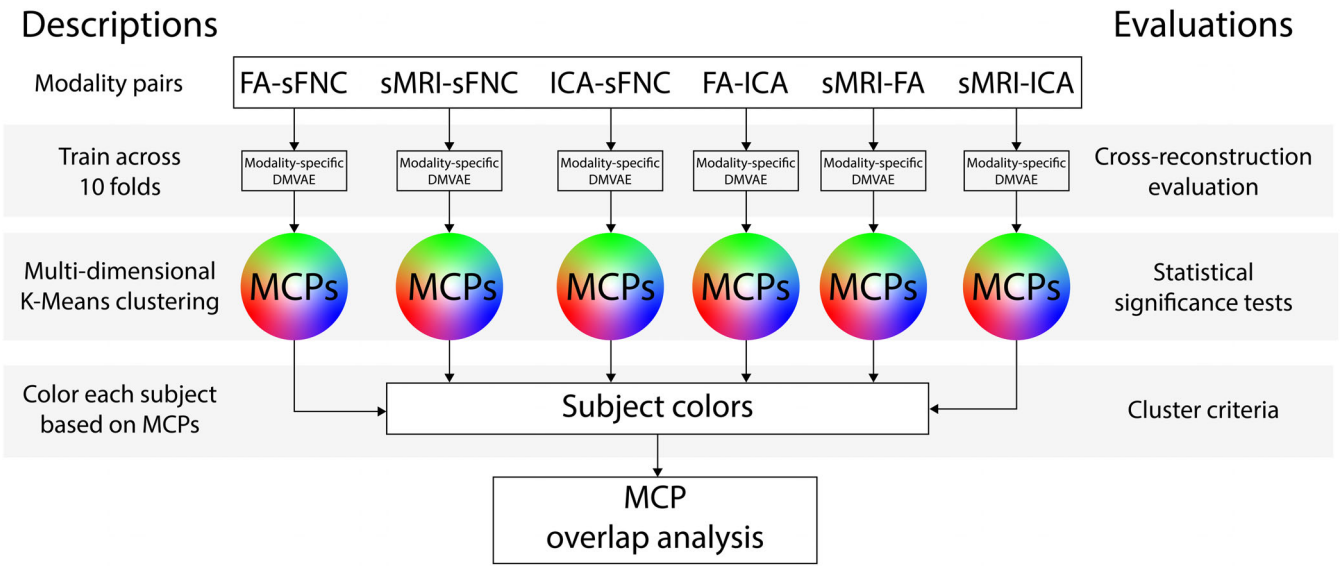

FIGURE 3 The order of analysis steps used in our chromatic fusion framework. This figure depicts the complete workflow of the analyses we perform in our framework and how they coincide. We start at the top with the modality pairs and work our way down to MCPs, and after that meta-MCPs.

TABLE 1 Data sample demographics.

	FA-sFNC	sMRI-sFNC	ICA-sFNC	FA-ICA	sMRI-FA	sMRI-ICA
N subjects	278	311	310	277	278	310
Patients (%)	49.28	48.55	48.39	49.10	49.28	48.39
Female (%)	25.54	26.05	26.13	25.63	25.54	26.13
Female patient (%)	23.36	23.84	24.00	23.53	23.36	24.0
Avg Age	38 ± 11	38 ± 11	38 ± 11	38 ± 11	38 ± 11	38 ± 11
Avg P length	17 ± 12	17 ± 11	17 ± 11	17 ± 12	17 ± 12	17 ± 11
AP (%)	89.78	88.74	88.67	89.71	89.78	88.67
AD (%)	37.23	37.09	36.67	36.76	37.23	36.67
Avg AP length	15 ± 11	15 ± 11	15 ± 11	15 ± 11	15 ± 11	15 ± 11
PANSS positive	15.5 ± 5.0	15.3 ± 5.0	15.2 ± 4.8	15.4 ± 4.8	15.5 ± 5.0	15.2 ± 4.8
PANSS negative	14.5 ± 5.7	14.3 ± 5.6	14.2 ± 5.6	14.4 ± 5.7	14.5 ± 5.7	14.2 ± 5.6
PANSS composite	−1.1 ± 6.5	−1.0 ± 6.3	−1.0 ± 6.3	−1.0 ± 6.5	−1.1 ± 6.5	−1.0 ± 6.3

Note: P in this table refers to psychosis, AP refers to anti-psychotic medication, and AD refers to anti-depressive medication. Thus, AP and AD in this table refer to the percentage of patients taking anti-psychotic and anti-depressive medication, respectively. PANSS is a symptom scale for schizophrenia. We show its positive, negative, and composite score.

San Francisco). Each consortium records diagnosis, age at the time of the scan, gender, illness duration, symptom scores, and current medication, when available. Furthermore, the inclusion criteria were that participants are between 18 and 65 years of age, and their schizophrenia diagnosis was confirmed by trained raters using the Structured Clinical Interview for DSM-IV (SCID) (First et al., 2002). All participants with a schizophrenia diagnosis were on a stable dose of antipsychotic medication either typical, atypical, or a combination for at least 2 months. Each participant with a schizophrenia diagnosis was clinically stable at the time of the scan. The control subjects were excluded based on current or past psychiatric illness based on the SCID assessment or in case a first-degree relative had an Axis-I psychotic disorder. Written informed consent was obtained from all study participants under protocols approved by the Institutional Review Boards at each consortium site.

2.7.2 | Preprocessing

We use structural MRI (sMRI), spatial ICA maps, and static functional network connectivity (sFNC) obtained through preprocessing with NeuroMark (Du et al., 2020). The sMRI volumes are preprocessed to voxel-based morphometry (VBM), whereas the spatial ICA maps and sFNC are both obtained by performing ICA on rs-fMRI data using the NeuroMark template. NeuroMark, using the NeuroMark_fMRI_1.0 template (template is released in the GIFT software at http://trendscenter.org/software/gift performs group ICA to extract 53 ICA components from the rs-fMRI signal and calculates the correlation between each to obtain the sFNC. For the spatial ICA components, we selected eight networks a priori to use in our model to reduce computational complexity, but use all 53 for the sFNC. Each spatial ICA component is used as a channel in a 3-dimensional convolutional neural network. Thus, handling more spatial ICA components requires more channels, and more complex training dynamics. The components we use for the spatial ICA components are the supplementary motor area, thalamus, middle inferior frontal gyrus, right inferior frontal gyrus, middle temporal gyrus, precentral gyrus, and inferior frontal gyrus. To obtain the voxel-based morphometry from sMRI data, the data are first processed with SPM 12 in a Matlab 2016 environment. The data are then further processed by segmenting it into modulated gray matter volumes (GMV) and smoothing those segmentations with a 6 mm FWHM Gaussian kernel.

For the FA maps, the diffusion-MRI (dMRI) scans were acquired on seven 3T Siemens Tim Trio System scanners and one 3T GE Discovery MR750 scanner at multiple sites. The scanning protocols are described in Keator et al. (2016)). The dMRI processing was then performed using FSL (www.fmrib.ox.ac.uk/fsl) and ANTs (Avants et al., 2009). The dMRI volumes were first corrected for eddy current distortions and head movement using the eddy (FSL 6.0) with advanced motion-induced signal dropout detection and replacement (Andersson & Sotiropoulos, 2016). Fractional anisotropy (FA) maps were then calculated from the diffusion tensor using dtifit (FSL). FA maps were normalized to the Montreal Neurologic Insititute (MNI) spaced FA template with a nonlinear registration by ANTs. Images with excessive motion, signal dropout, or noise were excluded from further analysis (Caprihan et al., 2011; Wu et al., 2015).

The FA maps and the sMRI volumes have different sizes, but we use the same convolutional architecture in the model for each. Namely, the FA maps were sampled at 1 mm and the sMRI volumes were sampled at 1.5 mm, so we resampled the FA maps to match the sMRI sampling using Scipy (Virtanen et al., 2020). The spatial ICA maps are sampled at 3 mm and cropped using a field-of-view (FOV), hence we use a separate convolutional architecture to handle the spatial ICA components, see Figure 1. We use ELUs (exponential linear units) (Clevert et al., 2015) as activations, GroupNorm (group normalization) (Wu & He, 2018) to stabilize the network and we add dropout

(Srivastava et al., 2014) after each layer to reduce overfitting. We also calculate a group mask for the FA and sMRI maps, to determine which voxels the model should reconstruct. We exclude values below 0.15 after rescaling between [0, 1] for the sMRI and FA data, and exclude values with an absolute value below 0.15 for the spatial ICA maps. Each of the modalities are z-scored based on the mean and standard deviation in the full dataset.

2.8 | Experiments

Chromatic fusion is performed on every combination of voxel-based morphometry (VBM), FA maps, static functional connectivity (sFNC), and spatial ICA maps. The model is trained with the Adam optimizer (Kingma & Ba, 2014) for 300 epochs with a learning rate of $1E-5$. A low learning rate is necessary to ensure stability during training. We use 16 latent dimensions for both private feature sets, and 32 dimensions for the shared features for each modality pair. After training a separate model for each modality pair, we identify the MCPs according to the algorithm described in Section 2.2 and evaluate the robustness of each MCP. We visualize the embedding space for each of the modality pairs, and then reconstruct a schizophrenia-enriched MCP in the original space of the modalities for each modality pair. To understand how each of the schizophrenia-enriched MCPs differ across modality pairs, we look at their overlap. A final experiment evaluates the performance of cross-reconstruction for each of the modality pairs on the test folds. The cross-reconstruction is compared to a reconstruction that has access to the private features of the modality (upper bound) and a reconstruction that is created using only the prior distribution (lower bound). To ensure reproducible results, we do not sample from the latent distributions, but rather take their mean because it is the most probable sample under a multivariate normal distribution.

3 | RESULTS

3.1 | MCP analysis

To quantify whether an MCP potentially represents neuropsychiatric and/or demographic factors we calculate the percentage of subjects in an MCP that also belong to a stratum. These percentages for sex and schizophrenia diagnosis are shown in Table 2.

The results in Table 2 show there is an interesting combination between unstable MCPs and more stable MCPs with fairly high percentages of subgroups in the population. All modality pairs produce at least one or two interesting MCPs, even if those MCPs are not as robust. To test whether the MCP pairs were significant in terms of the percentage of schizophrenia or female subjects captured by the MCP, we performed significance analyses, see Appendix A. Since our model captures non-linear relationships, we cannot simply regress out site effects beforehand, which is why we chose to assess post hoc if

TABLE 2 The meta-chromatic patterns for each modality pair.

MCPs	FA-sFNC	sMRI-sFNC	ICA-sFNC	FA-ICA	sMRI-FA	sMRI-ICA
0-R	72 ± 15	**82 ± 8**	93 ± 9	34 ± 16	71 ± 17	52 ± 18
0-SZ	53 ± 5	**72 ± 7**	62 ± 4	53 ± 11	29 ± 5	50 ± 11
0-F	28 ± 3	**24 ± 5**	24 ± 2	31 ± 6	42 ± 6	24 ± 6
1-R	61 ± 15	90 ± 3	95 ± 3	**63 ± 7**	72 ± 15	49 ± 14
1-SZ	53 ± 13	26 ± 2	24 ± 2	**73 ± 6**	48 ± 7	47 ± 10
1-F	26 ± 5	23 ± 1	27 ± 1	**25 ± 3**	19 ± 6	21 ± 7
2-R	84 ± 6	77 ± 12	**81 ± 7**	41 ± 16	**88 ± 5**	43 ± 14
2-SZ	27 ± 3	48 ± 9	**75 ± 3**	38 ± 12	**71 ± 2**	67 ± 6
2-F	21 ± 1	31 ± 5	**30 ± 1**	27 ± 3	**18 ± 3**	18 ± 7
3-R	**74 ± 14**	-	80 ± 20	82 ± 11	72 ± 11	71 ± 13
3-SZ	**73 ± 4**	-	31 ± 9	37 ± 8	55 ± 9	20 ± 4
3-F	**28 ± 3**	-	25 ± 1	12 ± 3	22 ± 7	38 ± 3
4-R	-	-	-	**92 ± 9**	-	**63 ± 17**
4-SZ	-	-	-	**83 ± 2**	-	**78 ± 5**
4-F	-	-	-	**18 ± 2**	-	**15 ± 6**
5-R	-	-	-	50 ± 10	-	46 ± 18
5-SZ	-	-	-	50 ± 7	-	44 ± 13
5-F	-	-	-	31 ± 3	-	34 ± 7

Note: We show the percentage-wise robustness (-R), subjects who are female (-F), and subjects diagnosed with schizophrenia (-SZ) for each MCP. The text of some schizophrenia-enriched MCPs (percentage above 70%) is made bold. Each cell's color indicates the color of the corresponding MCP, based on chromatic fusion. The color of each MCP is based on how far away from zero it is in the private feature set for the first and second modality (red and blue), and the shared feature set (green).

there were significant site effects for any of the MCPs with respect to the percentage of schizophrenia patients in a cluster. To do this, we calculated the standard deviation between the values shown in Table 2, and the median percentage of schizophrenia patients in a cluster over folds, and for each site. The standard deviation is less than 5% for all MCPs, and often around 1%–2%. This indicates that the percentage of patients in each MCP is generally stable across sites.

Every single modality pair also includes at least one MCP enriched for schizophrenia vs controls (more than 70% schizophrenia patients in the MCP), these MCPs are the upper quartile of schizophrenia-enriched clusters and are highlighted as bold in Table 2. Importantly, most of these schizophrenia-enriched clusters have low standard deviations across folds, and are often also robust. For example, for the ICA-sFNC pair, MCP 2 is highly robust (81%) with 75% schizophrenia subjects, and is green, meaning the irregularities are especially large in the shared features. For the FA-sFNC and sMRI-FA pairs, the most schizophrenia-enriched MCPs (3, and 2, respectively) are red-ish (pink) and blue, respectively. For the FA-sFNC pair, the FA maps are the first private features, and for the sMRI-FA pair, the FA maps constitute the second private features, so these MCPs both indicate irregularities in the FA maps for the schizophrenia-enriched MCPs. On the other hand, for FA-ICA and sMRI-ICA, cluster 4 contains irregularities in all feature sets, as indicated by the white color. It is notable that none of the schizophrenia-enriched MCPs are particularly skewed in terms of the number of male/female subjects, compared to the percentage of female subjects in the dataset, except for sMRI-FA MCP 0 (see Table 1).

The importance of the schizophrenia-enriched MCPs is also apparent from the significance analysis in Appendix A. The significant differences based on sex are often rather small or nonexistent, except for the sMRI-FA modality pair. MCP 3 for the FA-sFNC pair contains a corrected significant number of schizophrenia subjects compared to MCPs 0 ($p = 0.047$), and 2 ($p \leq 0.005$). For MCP 0 in the sMRI-sFNC pair, all of the comparisons are highly significant compared to the other clusters: MCP 1 ($p < 0.0005$) and MCP 2 ($p < 0.0005$). MCP 2 in the ICA-sFNC pair contains a significant number of schizophrenia compared to MCPs 1 ($p < 0.0005$) and 3 ($p < 0.0005$), but not 0. This is because MCP 0 also contains a relatively large number of schizophrenia subjects (62%), which is significant compared to MCPs 1 ($p < 0.0005$) and 3 ($p < 0.0005$). Both FA-ICA MCP 1 and 4 are enriched for schizophrenia subjects, and are significant with respect to clusters 2 ($p < 0.0005$ and $p = 0.004$) and 3 ($p = 0.007$ and $p = 0.019$). The significance values for MCP 4 are lower due to fewer subjects in MCP 4, but the MCP itself is robust across folds 93%, and contains 84% schizophrenia subjects. For the sMRI-FA pair, MCP 2 contains a significant number of schizophrenia subjects compared to cluster 0 ($p < 0.0005$). MCP 0 for this modality pair actually contains a significantly small number of schizophrenia subjects compared to all other MCPs: 1 ($p < 0.001$), 2 ($p < 0.0005$), and 3 ($p < 0.0005$). For the sMRI-ICA pair, MCP 4 contains a significant number of schizophrenia subjects compared to MCPs 0 ($p = 0.018$), 1 ($p = 0.001$), 3 ($p < 0.0005$), and 5 ($p < 0.0005$).

3.2 | Chromatic fusion space shapes

We visualize the embedding space for each modality pair in Figure 4. The colors of the MCPs and subjects in the figure are based on the L2-norm from a subject to its assigned cluster, as explained in Section 2.2. To visualize the 64-dimensional subjects and clusters, the mean of the distributions are visualized in a 2-dimensional space using t-SNE (van der Maaten & Hinton, 2008). Namely, t-SNE is a commonly used method to visualize spaces in 2-dimensions.

Each modality pair clearly has unique color patterns and structures in the embedding space, shown in Figure 4. Furthermore, this visualization highlights how MCPs roughly compare to each other spatially in the embedded chromatic space. It is clear from this view that MCP 4 in the FA-ICA is a small MCP with very irregular subjects (white color), seemingly far away from the rest of the MCPs. This leads to very dark colors for the other MCPs because their irregularities are not as extreme as this MCP. In this case, we have put the MCPs in the color wheel legend near the color that they are closest to perceptually and numerically since the color wheel does not contain dark colors. Additionally, for MCP 2 and MCP 1 the FA-sFNC and sMRI-sFNC modality pairs, respectively, seem to be further away from the other MCPs. Neither are schizophrenia-enriched MCPs, but rather control-enriched MCPs. For the sMRI-ICA modality pair, however, it is harder to make clear distinctions, which is also reflected in the colors of the MCPs and how chromatic the space is. For this modality pair, most MCPs have similar colors, and as a result, their subjects are also colored a reddish green.

3.3 | Visualizing schizophrenia-enriched MCPs

To qualitatively understand what types of irregularities the main schizophrenia-enriched MCPs represent, we have reconstructed each MCP using the decoder in our model. Instead of using the MCP center however, we take the schizophrenia patients assigned to each schizophrenia-enriched MCP and reconstruct their average in the latent space. This allows us to directly look at the average of the schizophrenia subjects, without adding heterogeneity from control subjects. Figure 5 shows these reconstructions for each modality pair.

The reconstructions of the MCPs in Figure 5 show interesting patterns. Mainly, we observe differences between MCPs that have the same modalities. For instance, the sMRI patterns in the schizophrenia-enriched MCP for the sMRI-sFNC modality pair has opposite directionality in some areas compared to the schizophrenia-enriched sMRI-FA and sMRI-ICA pairs. A similar pattern of opposite directionality occurs for the spatial ICA patterns for the ICA-sFNC MCP, when compared with FA-ICA and sMRI-ICA. To test our hypothesis that these MCPs capture different subgroups of schizophrenia patients, we perform a heterogeneity analysis on the schizophrenia-enriched MCPs.

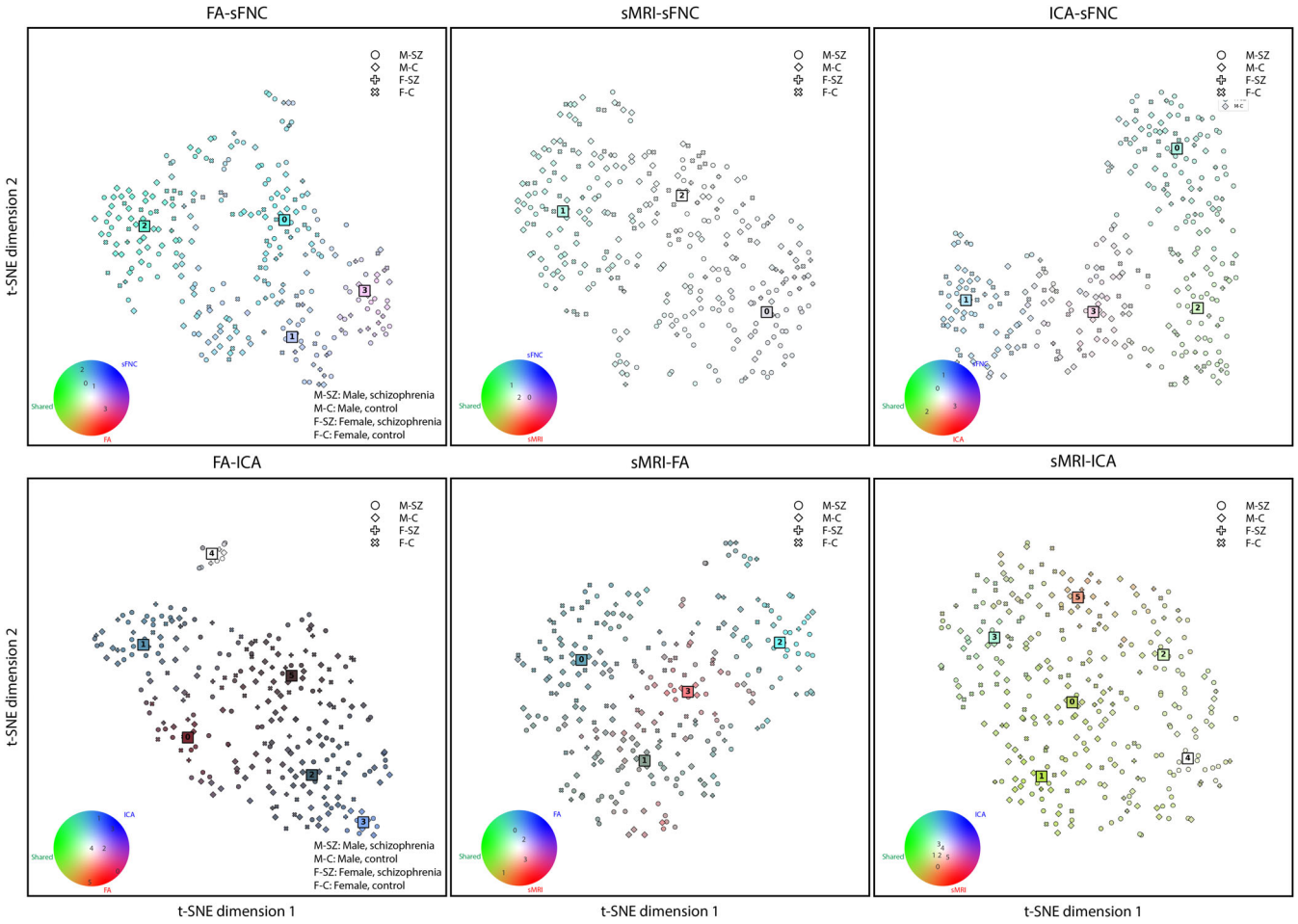

FIGURE 4 A 2D t-SNE plot of the embedding space for each of the modality pairs. This is a visualization of what the embeddings for each subject look like, they have been chromatically colored according to our framework. The color wheels in the bottom left of each subfigure show the location of MCPs on the color wheel in terms of their irregularities in the shared and private feature sets.

3.4 | Heterogeneity of schizophrenia as a function of modality pair

The fact that different modality pairs lead to schizophrenia-enriched clusters with different and sometimes even opposing modality-based differences ties into our next results. Namely, how many schizophrenia subjects are distinctly captured by each of the schizophrenia-enriched MCPs. For this analysis, we use all schizophrenia-enriched MCPs (%SZ >70): FA-sFNC-MCP-3, sMRI-sFNC-MCP-0, ICA-sFNC-MCP-2, FA-ICA-MCP-1, FA-ICA-MCP-4, sMRI-FA-MCP-2, and sMRI-ICA-MCP-4. On average across folds, these clusters account for 67% ± 2 of all schizophrenia subjects in the dataset. The results in Figure 6 show the average percentage of schizophrenia subjects captured by the row-wise MCP that are not captured by the column-wise MCP, divided by the number of schizophrenia subjects in the row-wise MCP. This is thus a metric of distinct schizophrenia subjects captured by the row-wise MCP, and a higher percentage means the MCP captures a higher percentage of distinct schizophrenia subjects.

The most distinct aspect of Figure 6 is that FA-ICA MCP 4 is almost entirely made up of unique subjects compared to all other schizophrenia-enriched MCPs. This is a relatively small MCP, so it only captures a highly distinct group of schizophrenia subjects. In fact, in our analyses of the cognitive and symptom scores in Appendix D, we show that this MCP corresponds to the highest average PANSS positive symptom scores compared to other clusters. Generally, many of these MCPs contain distinct schizophrenia subjects when compared with MCPs from other modalities. Many MCPs contain atleast 40%–50% unique schizophrenia subjects compared to other MCPs. This potentially indicates that heterogeneity, and what schizophrenia subjects are highlighted, is a function of what modalities are paired.

3.5 | The importance of the shared features

We perform a more in-depth analysis of the robustness across folds and correlation of each latent dimension in the shared features to schizophrenia in Appendix C. Based on the most robust shared latent dimensions, we select two latent dimensions to highlight how interpolating from one end of the dimension to the other leads to covariations of the modality pair. For the following latent dimensions these

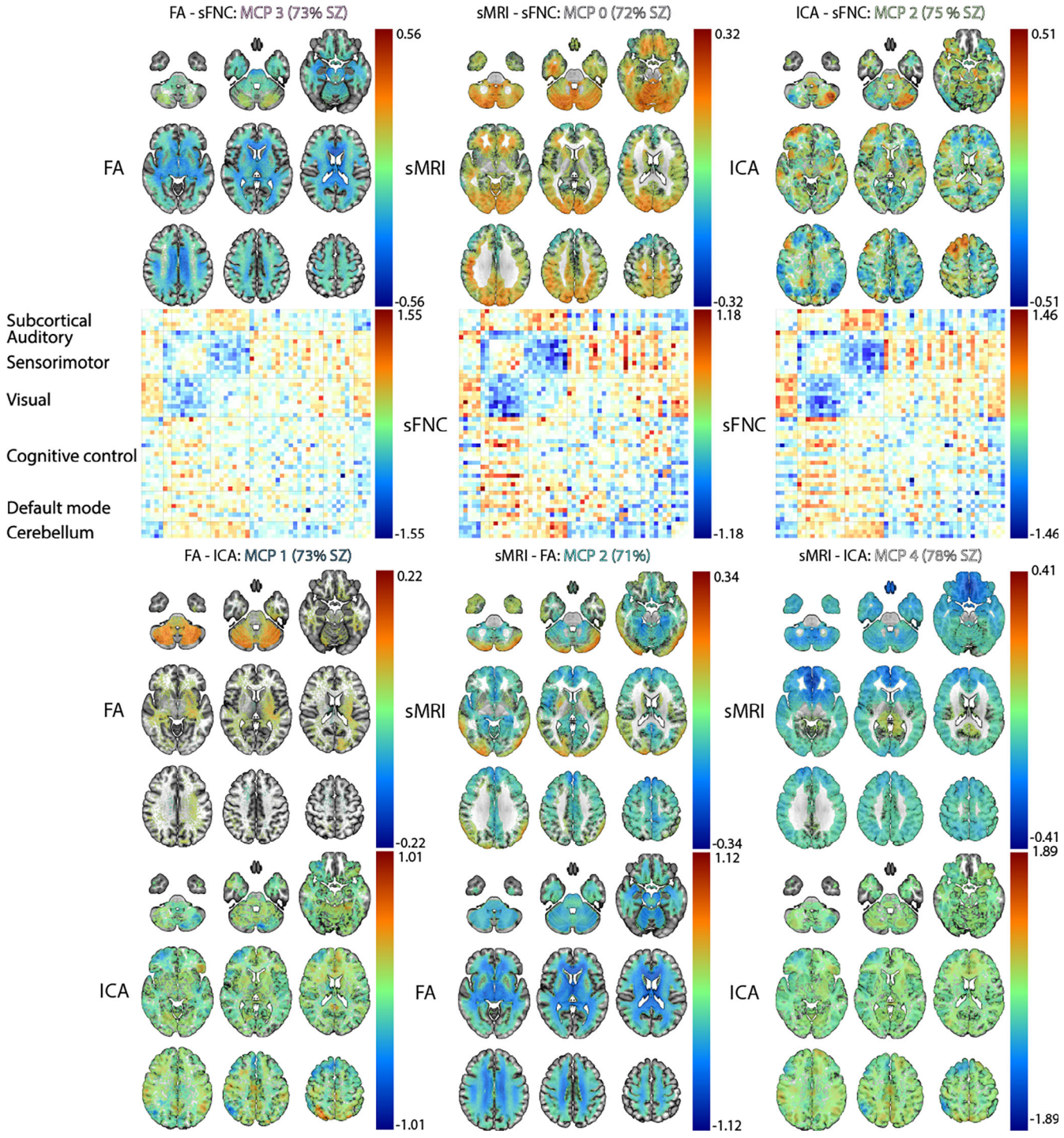

FIGURE 5 The reconstruction of schizophrenia-enriched MCPs. The percentages next to the name of the MCP indicate the average percentage of schizophrenia subjects in the cluster across folds. Note that the values for the color bar differ for each modality pair, this is to ensure there is enough contrast to see differences in positive and negative changes.

interpolations are aligned with schizophrenia diagnoses and are shown in Figure 7.

Figure 7 indicates that in the FA-sFNC pair, as we interpolate from schizophrenia subjects to control subjects, we see an increase in modularity and connectivity between the subcortical and sensorimotor regions, and a decrease in self-connectivity for the visual regions. This is coupled with higher fractional anisotropy in the cerebellum for schizophrenia subjects that non-linearly interpolates into a decrease in fractional anisotropy in a more anterior part of the cerebellum. Since these are shared features, it is interesting that the interpolations for both modalities follow a gradient for the cerebellum. For the sMRI-sFNC pair, we almost see a split in the scatter plot at the bottom

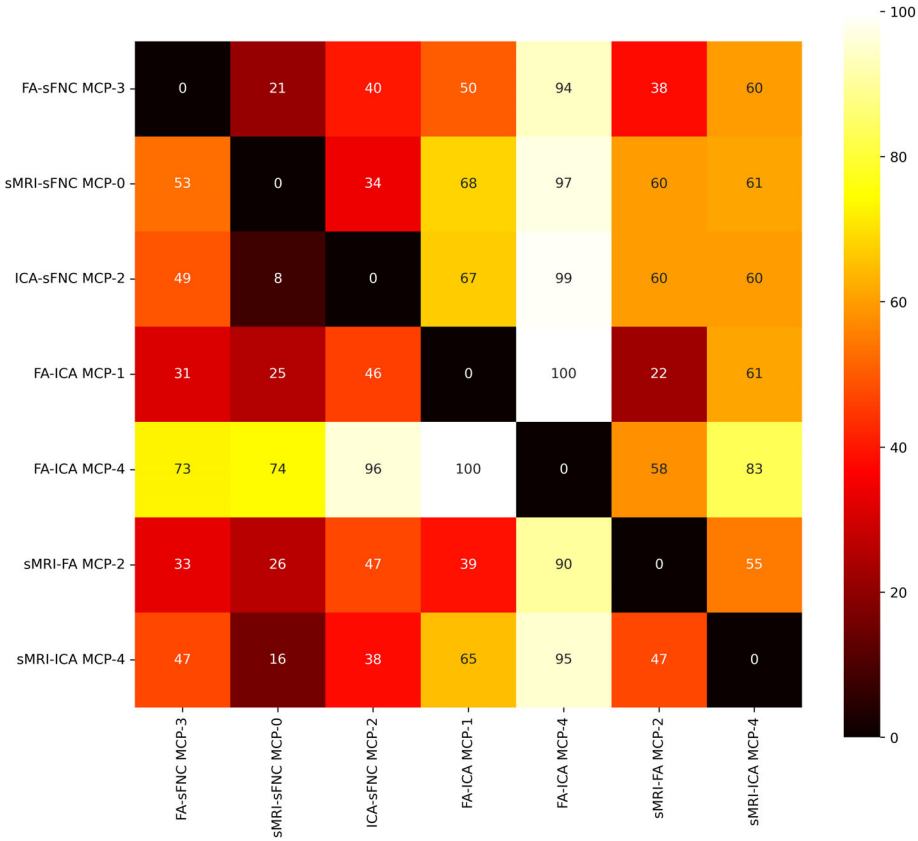

FIGURE 6 The distinct percentage of schizophrenia subjects each of the schizophrenia-enriched MCPs uniquely capture compared to each other. The unique schizophrenia subjects in the row MCP compared to the column MCP are divided by the total number of schizophrenia subjects in the row MCP. Higher values correspond to a higher percentage of unique schizophrenia subjects captured by that row MCP compared to the column MCP.

of Figure 7 (around the fourth panel from the left) between schizophrenia patients with higher and lower general voxel-based morphometry. Although the voxel-based morphometry in the cerebellum for more schizophrenia-enriched patients interpolates to a decrease in voxel-based morphometry and moves more anterior, the rest of the brain moves from lower voxel-based morphometry to increased voxel-based morphometry on the control-enriched side. This is coupled with an increase in the modularity of the sFNC, but also an increase in the connection between sensorimotor and visual regions for control subjects.

3.6 | Cross-reconstruction

To evaluate the shared features with respect to how well it can help reconstruct other modalities, we compare the performance of the cross-reconstructions across all the modality pairs. The metric we use to compare the cross-reconstruction with two baselines is the mean squared error of the reconstruction. The difference is evaluated on the test set and averaged across all 10 folds, and the results are shown in Table 3. Cross-reconstruction is evaluated for two baselines. The first baseline uses the mean of the prior (zero) to reconstruct the "missing" modality. The second baseline is called normal and uses the private and shared features for a modality to reconstruct it. The cross-reconstruction uses the shared features of the other modality and the mean of the prior (zero) for the private features to reconstruct the "missing" modality.

Table 3 shows that the cross-reconstruction of modalities generally improves with the shared information. Although the results differ in terms of the magnitude of the improvement, all of these shared features have a fairly high KL-divergence, which indicates that they contain information about both modalities. Thus, these results indicate that the shared features generally help reconstruction, and sometimes come close to knowing both the private and shared features of the other modality. How much shared features help with reconstructions, however, depends on the modality pair.

4 | DISCUSSION

In this work, we present an intuitive and flexible framework to facilitate new insights into multimodal neuroimaging data. By representing information from modalities as colors, we can intuitively visualize private and shared feature sets for modality pairs. Our method identifies schizophrenia-enriched clusters for each modality pair. The clusters are assigned colors to indicate important meta-chromatic patterns (MCPs) and define a chromatic space. Although these MCPs overlap across modality pairs in terms of what subjects are captured by schizophrenia-enriched MCPs, we find that different modality pairs highlight distinct subgroups of schizophrenia subjects. Together with visualizations and statistical results for the schizophrenia-enriched MCPs, and interpolations of the shared dimensions for certain modality pairs, we show how critical multi-modal analyses are to further our understanding of spectrum psychiatric disorders, such as schizophrenia.

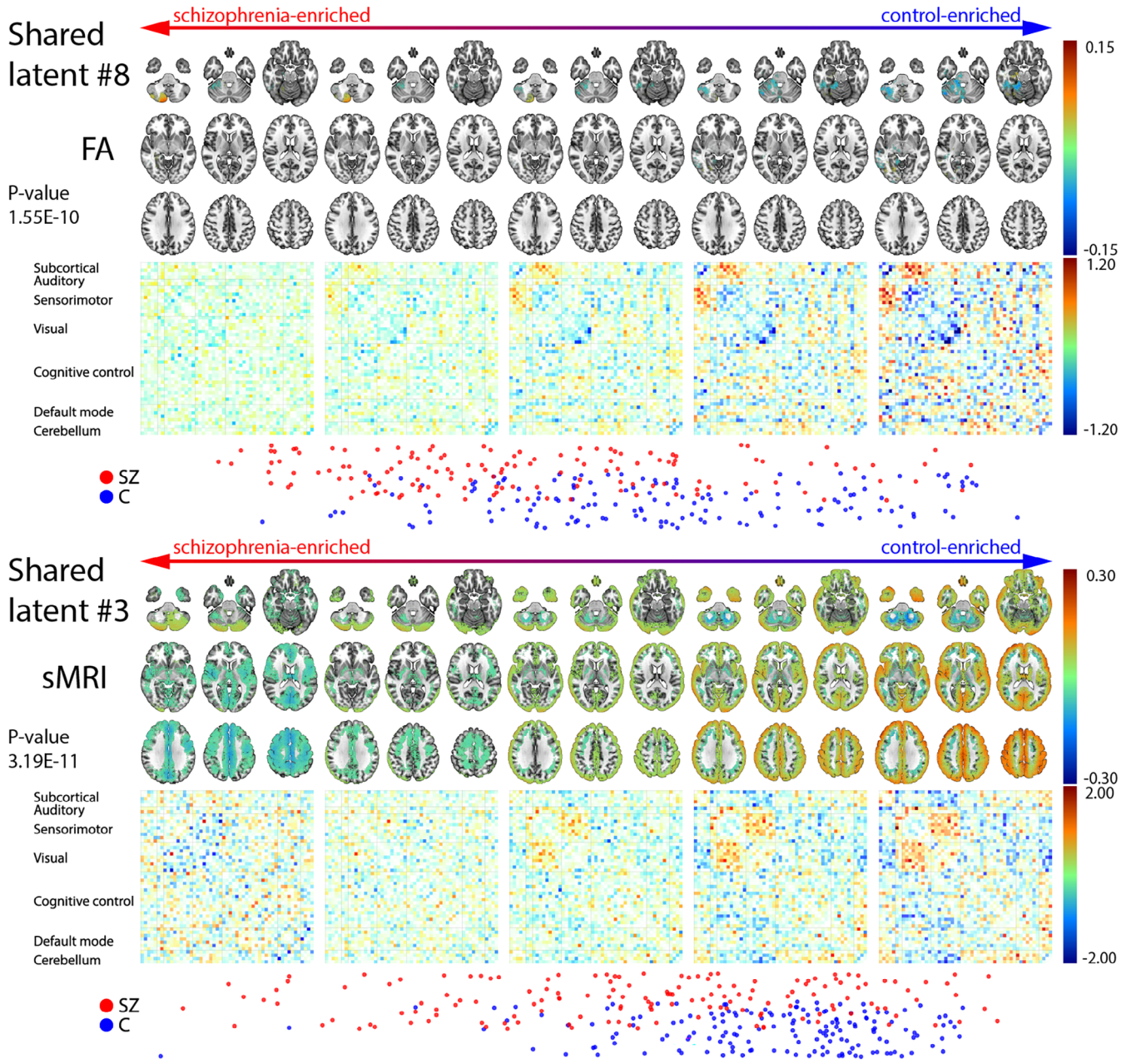

FIGURE 7 Interpolations for two representative latent dimensions from schizophrenia-enriched to control-enriched. The two latent dimensions are taken from the FA-sFNC, and sMRI-sFNC modality pairs, and show how the modalities change along the latent dimension, from a more schizophrenia-enriched part of the latent dimension to a control-enriched part of the latent dimension. We have included a plot of the schizophrenia and control subjects, and the p-value for fold 0 in the figure.

Our analyses are centered around a framework that views the contributions of modalities in modality pairs as colors. Especially as we move to the inclusion of more modalities, the visualization, and interpretation of combinations of modalities are more naturally done by considering them as different colored lights that chromatically fuse into a certain perspective on diseases or demographic variables. Furthermore, to keep in line with the RDoC initiative, we propose a framework that allows multiple modalities or units of analysis to be included in a single framework that aims to study a mental disorder. We refrain from predicting binary labels and only use unsupervised methods to study schizophrenia. We find that the schizophrenia-enriched MCPs capture more than 67% of the schizophrenia subjects in our sample. This means that along either of the modality pairs, these schizophrenia subjects show clear deviations from the rest of the sample, which is subsequently captured in these schizophrenia-enriched MCPs. To dive deeper into what these specific deviations mean in the original space of the modality, we visualize the most important schizophrenia-enriched MCP for each modality pair, revealing results that align with previous work, but also extend it in an interesting, and understudied multimodal way.

TABLE 3 The mean squared error for cross-reconstructed modalities.

Modalities	Prior	Normal	Cross
sMRI→FA	0.9902 ± 0.2155	0.6814 ± 0.1278	0.7108 ± 0.1411
FA→sMRI	0.9717 ± 0.0451	0.8681 ± 0.0344	0.9219 ± 0.0444
FA→sFNC	0.9717 ± 0.0451	0.8761 ± 0.0357	0.9113 ± 0.0369
sFNC→FA	1.0245 ± 0.0460	0.8197 ± 0.0324	1.0152 ± 0.0431
sMRI→sFNC	0.9874 ± 0.1635	0.7012 ± 0.0942	0.7683 ± 0.1159
sFNC→sMRI	1.0198 ± 0.0659	0.8091 ± 0.0464	1.0022 ± 0.0624
ICA→sFNC	1.0000 ± 0.0056	0.9960 ± 0.0056	0.9975 ± 0.0057
sFNC→ICA	1.0239 ± 0.0532	0.7995 ± 0.0308	0.9212 ± 0.0417
FA→ICA	0.9709 ± 0.0828	0.8646 ± 0.0433	0.8706 ± 0.0459
ICA→FA	1.0001 ± 0.0067	0.9983 ± 0.0073	1.0012 ± 0.0060
sMRI→ICA	0.9878 ± 0.1252	0.6747 ± 0.0727	0.6938 ± 0.0798
ICA→sMRI	1.0000 ± 0.0056	0.9972 ± 0.0057	1.0000 ± 0.0055

Note: The mean squared error is calculated between the voxels of the reconstructed volumes and the ground truth volumes.

Our findings in Figure 5 regarding schizophrenia-enriched meta-chromatic patterns in Figure 5 show some general trends. First, each of the schizophrenia-enriched MCPs for the modality pairs that contain sFNC in its pair, show hypoconnectivity between visual and sensorimotor regions and especially for the ICA-sFNC pair hyperconnectivity between visual-cerebellum and visual-subcortical regions. This is a general trend in each of the MCPs, although which components are most hypo-connected and how hyperconnected the aforementioned regions are differs for each of the modality pairs. Hypoconnectivity between the visual and sensorimotor cortex for schizophrenia subjects has previously been linked to schizophrenia (Chen et al., 2015). Reduced connectivity between these areas may have an impact on self-processing in patients (Chen et al., 2015) and generally be related to early-stage visual processing deficits in schizophrenia (Butler & Javitt, 2005). Furthermore, subcortical-visual and cerebellum-visual hyperconnectivity aligns with previous unimodal work within this cohort (Damaraju et al., 2014; Ford et al., 2015) and in other cohorts (Liang et al., 2006). These connectivity patterns are paired with reduced fractional anisotropy near the corpus callosum and regions superior to the corpus callosum in the FA-sFNC pair. Notably, a decrease in FA strength in the cerebellum, corpus callosum, and superior longitudinal fasciculi have previously been linked to schizophrenia (Koch et al., 2010; Shergill et al., 2007). The decrease in fractional anisotropy especially near the corpus callosum is potentially related to schizophrenia-like psychosis due to structural defects (Koch et al., 2010; Walterfang et al., 2005). We see similar patterns of reduced FA strength in the sMRI-FA pair, where decreased FA is also more prominent in the cerebellum. Interestingly, the FA-ICA pair shows increased FA strength in the cerebellum, although only slightly. This potentially reflects a slightly distinct schizophrenia subgroup without the reduced FA strength in the cerebellum. For the sMRI-sFNC pair, the connectivity pattern for the sFNC are linked with increased voxel-based morphometry most pronounced in the cerebellum, occipital lobe, and near the motor areas, with some reduced voxel-based morphometry in the superior frontal lobe. For all other sMRI-based MCPs, we generally see reductions in frontal lobe voxel-based morphometry, which is in line with previous work (Fornito et al., 2009). In a similar opposite directional pattern, the sMRI-sFNC MCP may capture some schizophrenia subjects with different patterns from the other sMRI-based MCPs, accentuating schizophrenia's heterogeneity. We also observe a different pattern for the ICA-based MCPs, where the ICA-sFNC shows increased spatial ICA map strength in the left frontal lobe and directional asymmetry in the cerebellum (decrease on the left and increase on the right). However, for the FA-ICA and sMRI-ICA MCPs, the spatial ICA maps have decreased spatial ICA map strengths in the left frontal lobe. Together, the results in Figures 5 and 6 clearly indicate that our method can find interesting schizophrenia subgroups, indicating that heterogeneity may be a function of what modalities are paired during training.

An important aspect of schizophrenia we highlight in this work is how the meta-chromatic patterns we find are associated with different subgroups of schizophrenia patients. By analyzing a wider range of modalities, and specifically training our model to find interpretable feature sets between pairs of modalities, we are able to understand how these subgroups change with different modality pairs. Although the subgroups are not unique for each modality pair, it is clear that additional modalities provide us with information about subjects that may not have been captured by a schizophrenia-enriched cluster if we had only considered one or two modalities. This also speaks to the heterogeneous nature of schizophrenia, where some subjects may deviate across most brain measures, and some subjects only deviate along specific brain measures or the shared information in specific brain measures. By highlighting schizophrenia from multiple directions, and conceptualizing an intuitive framework around this type of analysis, we also specifically follow the trans-diagnostic NIMH research domain criteria (RDoC) initiative (Sanislow et al., 2019).

We additionally emphasize an important finding regarding spectrum mental disorders: the significance of extracting shared features from a pair of modalities. In Appendix C, we demonstrate that for each modality pair, there exist replicable shared latent dimensions. Furthermore, we observe that at least a few of these latent dimensions exhibit high correlations to schizophrenia. The reason that these

latent dimensions are of particular interest is that we can easily interpolate between ends of the dimension that are either control-enriched or schizophrenia-enriched. We also find that the interpolations are largely non-linear and could not have been captured by a linear model. These interpolations help us view certain normative imaging deviations related to schizophrenia on a spectrum as well. There is a wealth of information that is encoded exactly on the border between schizophrenia-enriched and control-enriched shared latent dimensions, namely how the extracted shared features transition from schizophrenia to controls. Specifically, the significance of the correlations to schizophrenia as shown in Figure 7 highlight its potential for future analyses. The patterns we see change along two highlighted latent dimensions show distinct modality-specific patterns exist at the extreme end of the interpolative patterns. Furthermore, we see evidence of reduced modular functional patterns on the schizophrenia-enriched side of the latent dimension, as linked to schizophrenia previously (Yu et al., 2012), and focal spatial regions, for the FA-sFNC pair specifically. Notably, the patterns we obtain for sFNC, which is part of both modality pairs, is different for both the highlighted FA-sFNC and sMRI-sFNC latent dimension. This again indicates how coupling different modalities can result in distinct patterns being uncovered by multimodal machine learning models.

4.1 | Limitations

The validity of meta-chromatic patterns is partly dependent on the quality of the reconstructions that the DMVAE model produces. It is therefore important to improve the reconstructions by adapting the model or architectures to neuroimaging modalities and improve the cross-reconstructions between the modalities. Furthermore, not all of the MCPs are as robust as others, which is likely due to the stochastic nature of training a variational autoencoder and the differences in distribution between each of the folds. This can be improved by imposing more inductive biases into the architecture and model, or by using larger datasets. This leads to another point, namely that this framework needs to be tested on more datasets to show its robustness across datasets.

It is challenging to address site effects in the context of machine learning models that can capture complex and nonlinear relationships. To evaluate the potential impact of the acquisition site on the result we calculated the standard deviation between the median (across folds) percentage of patients from each site in a cluster. Results showed that the percentage of patients in a cluster was always within 5% of the mean. The patient/control ratio was thus consistent across sites. This gives us confidence in the robustness of the results to site effects. However, for larger datasets analyzed in future work, we plan to continue to evaluate this issue and develop additional metrics that can quantify site effects.

Other than site effects, schizophrenia clusters could also be proxies for uncontrolled confounders, such as severe mental illness, prolonged exposure to psychotropic drugs, and lower socioeconomic status. This is a more general issue with data-driven schizophrenia analysis and should be carefully assessed in exploratory studies. Some modalities may also be more compatible with each other than others. For example, modalities with largely mutually exclusive information can lead to very noisy shared feature sets because the model cannot find any shared features. Thus, it is important to be careful when applying these models to any two modalities.

4.2 | Future work

The framework can be expanded to additional datasets and more modalities. Future work can also extend the model and MCP framework to move beyond pair-wise to N-way unique and shared links among modalities. Additionally, there is a potential to discover more entangled shared features by learning representations from minimally pre-processed rs-fMRI directly, together with the FA maps and structural MRI volumes. The incorporation of rs-fMRI as a timeseries, rather than static FNC will facilitate additional insights as in parallel group ICA + ICA (Qi et al., 2019) and also allow for the fusion of other dynamic modalities, such as EEG. Learning representations from minimally preprocessed modalities can be coupled with ingenious ways to incorporate inductive biases such as group differentiation, regularizations on the weights in the network, or constraints on the latent features such as sparsity. Another integration into these multimodal frameworks is to condition them on cognitive or symptom scores, such that we can leverage information-rich cognitive measures that are not binary. There are thus wide-ranging possibilities to increase the utility and generality of this framework. Furthermore, the sub-types we find in this work and the method we propose need to be validated in other (schizophrenia) datasets to improve our understanding of both the mental disorders under study and the importance of multimodal methodologies. One way in which we can do this is evaluate our model on other samples of schizophrenia subjects, or on larger pooled samples. Furthermore, digging deeper into specific subgroups and untangling why their patterns are different from other sub-groups under specific multimodal pairings can bring us a step closer to understanding heterogeneity in schizophrenia.

5 | CONCLUSION

The presented framework can be used in various ways, both as an exploratory tool, to perform hypothesis-testing, to evaluate the enrichment of modalities with each other and specific inter-modality patterns hypothesized apriori, or to evaluate the heterogeneity of schizophrenia subjects within a sample. Our model can visualize individualized inter-modal relationships, a novel aspect of our work relevant to many applications, including brain development/aging, clinical studies, etc. To conclude, we have shown this framework mostly as an exploratory tool but linked our findings back to previous work on schizophrenia or on this sample specifically. In doing this, we showed how training an autoencoder to find private and shared features from a pair of modalities leads to meta-chromatic patterns that are schizophrenia-enriched, each modality pair with a different schizophrenia subgroup. We also take a first step toward understanding and

evaluating the heterogeneity among schizophrenia subjects in our sample and visualize interesting inter-modality patterns and interpolations. For instance, we observe a decrease in the modularity of functional connectivity and decreased visual-sensorimotor connectivity for schizophrenia patients with the FA-sFNC and sMRI-sFNC modality pairs, respectively. The visual-sensorimotor hypoconnectivity may indicate impaired self-processing in patients. Additionally, our results generally indicate decreased fractional corpus callosum anisotropy, which is linked to psychosis, and decreased spatial ICA map and voxel-based morphometry strength in the superior frontal lobe for patients across multiple modality pairs.

ACKNOWLEDGMENTS

This material is supported by the National Science Foundation under Grant No. 2112455 and the National Institutes of Health Grant #R01EB006841. Eloy P.T. Geenjaar was supported by the Georgia Tech/Emory NIH/NIBIB Training Program in Computational Neuralengineering (T32EB025816).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ORCID

Eloy P.T. Geenjaar https://orcid.org/0000-0001-5448-6358
Noah L. Lewis https://orcid.org/0000-0002-5712-2434
Lei Wu https://orcid.org/0000-0002-9705-1829

REFERENCES

Andersson, J. L., & Sotiropoulos, S. N. (2016). An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging. *NeuroImage*, *125*, 1063–1078.

Avants, B. B., Tustison, N., & Song, G. (2009). Advanced normalization tools (ANTS). *Insight j*, *2*(365), 1–35.

Baltrušaitis, T., Ahuja, C., & Morency, L. P. (2018). Multimodal machine learning: A survey and taxonomy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *41*(2), 423–443.

Bhugra, D., Tasman, A., Pathare, S., Priebe, S., Smith, S., Torous, J., Arbuckle, M. R., Langford, A., Alarcón, R. D., Chiu, H. F. K., First, M. B., Kay, J., Sunkel, C., Thapar, A., Udomratn, P., Baingana, F. K., Kestel, D., Ng, R. M. K., Patel, A., … Ventriglio, A. (2017). The WPA-lancet psychiatry commission on the future of psychiatry. *The Lancet Psychiatry*, *4*(10), 775–818.

Butler, P. D., & Javitt, D. C. (2005). Early-stage visual processing deficits in schizophrenia. *Current Opinion in Psychiatry*, *18*(2), 151–157.

Calhoun, V. D., & Sui, J. (2016). Multimodal fusion of brain imaging data: A key to finding the missing link (s) in complex mental illness. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *1*(3), 230–244.

Calhoun, V. D., Liu, J., & Adal, T. (2009). A review of group ICA for fMRI data and ICA for joint inference of imaging, genetic, and ERP data. *NeuroImage*, *45*(1), S163–S172.

Caprihan, A., Abbott, C., Yamamoto, J., Pearlson, G., Perrone-Bizzozero, N., Sui, J., & Calhoun, V. D. (2011). Source-based morphometry analysis of group differences in fractional anisotropy in schizophrenia. *Brain Connectivity*, *1*(2), 133–145.

Chen, X., Duan, M., Xie, Q., Lai, Y., Dong, L., Cao, W., Yao, D., & Luo, C. (2015). Functional disconnection between the visual cortex and the sensorimotor cortex suggests a potential mechanism for self-disorder in schizophrenia. *Schizophrenia Research*, *166*(1–3), 151–157.

Clevert, D. A., Unterthiner, T., & Hochreiter, S. (2015). Fast and accurate deep network learning by exponential linear units (elus). arXiv Preprint arXiv:151107289.

Damaraju, E., Allen, E. A., Belger, A., Ford, J. M., McEwen, S., Mathalon, D., Mueller, B. A., Pearlson, G. D., Potkin, S. G., Preda, A., Turner, J. A., Vaidya, J. G., van Erp, T. G., & Calhoun, V. D. (2014). Dynamic functional connectivity analysis reveals transient states of dysconnectivity in schizophrenia. *NeuroImage: Clinical*, *5*, 298–308.

du, Y., Fu, Z., Sui, J., Gao, S., Xing, Y., Lin, D., Salman, M., Abrol, A., Rahaman, M. A., Chen, J., Hong, L. E., Kochunov, P., Osuch, E. A., Calhoun, V. D., & Alzheimer's Disease Neuroimaging Initiative. (2020). NeuroMark: An automated and adaptive ICA based pipeline to identify reproducible fMRI markers of brain disorders. *NeuroImage: Clinical*, *28*, 102375.

Fedorov, A., Geenjaar, E., Wu, L., DeRamus, T. P., Calhoun, V. D., & Plis, S. M. (2021). Tasting the cake: Evaluating self-supervised generalization on out-of-distribution multimodal MRI data. arXiv preprint arXiv:210315914.

Fedorov, A., Sylvain, T., Geenjaar, E., Luck, M., Wu, L., DeRamus, T. P., Kirilin, A., Bleklov, D., Calhoun, V., & Plis, S. (2021). Self-supervised multimodal domino: In search of biomarkers for alzheimer's disease. In *2021 IEEE 9th International Conference on Health Informatics (ICHI)* (pp. 23-30). IEEEE.

First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. (2002). *Structured clinical interview for DSM-IV-TR axis I disorders, research version* (research ed.). SCID-I/P.

Ford, J. M., Palzes, V. A., Roach, B. J., Potkin, S. G., van Erp, T. G., Turner, J. A., Mueller, B. A., Calhoun, V. D., Voyvodic, J., Belger, A., Bustillo, J., Vaidya, J. G., Preda, A., McEwen, S. C., Functional Imaging Biomedical Informatics Research Network, & Mathalon, D. H. (2015). Visual hallucinations are associated with hyperconnectivity between the amygdala and visual cortex in people with a diagnosis of schizophrenia. *Schizophrenia Bulletin*, *41*(1), 223–232.

Fornito, A., Yücel, M., Patti, J., Wood, S. J., & Pantelis, C. (2009). Mapping grey matter reductions in schizophrenia: An anatomical likelihood estimation analysis of voxel-based morphometry studies. *Schizophrenia Research*, *108*(1–3), 104–113.

Garrity, A. G., Pearlson, G. D., McKiernan, K., Lloyd, D., Kiehl, K. A., & Calhoun, V. D. (2007). Aberrant "default mode" functional connectivity in schizophrenia. *American Journal of Psychiatry*, *164*(3), 450–457.

Hinton, G. E. (1999). Products of experts. In *1999 Ninth International Conference on artificial neural networks ICANN 99* (Vol. 1, pp. 1–6). UToronto.

Hu, D., Zhang, H., Wu, Z., Wang, F., Wang, L., Smith, J. K., Lin, W., Li, G., & Shen, D. (2020). Disentangled-multimodal adversarial autoencoder: Application to infant age prediction with incomplete multimodal Neuroimages. *IEEE Transactions on Medical Imaging*, *39*(12), 4137–4149.

Hutchison, R. M., Womelsdorf, T., Allen, E. A., Bandettini, P. A., Calhoun, V. D., Corbetta, M., Della Penna, S., Duyn, J. H., Glover, G. H., Gonzalez-Castillo, J., Handwerker, D. A., Keilholz, S., Kiviniemi, V., Leopold, D. A., de Pasquale, F., Sporns, O., Walter, M., & Chang, C. (2013). Dynamic functional connectivity: Promise, issues, and interpretations. *NeuroImage*, *80*, 360–378.

Keator, D. B., van Erp, T. G., Turner, J. A., Glover, G. H., Mueller, B. A., Liu, T. T., Voyvodic, J. T., Rasmussen, J., Calhoun, V. D., Lee, H. J., Toga, A. W., McEwen, S., Ford, J. M., Mathalon, D. H., Diaz, M., O'Leary, D. S., Jeremy Bockholt, H., Gadde, S., Preda, A., … FBIRN. (2016). The function biomedical informatics research network data repository. *NeuroImage*, *124*, 1074–1079.

Kim, T., Eltoft, T., & Lee, T. W. (2006). *Independent vector analysis: An extension of ICA to multivariate components. In: Independent component*

analysis and blind signal separation: 6th international conference, ICA 2006 (pp. 165–172). March 5–8, 2006. Proceedings 6 Springer.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization (version 9). *arXiv*. https://doi.org/10.48550/ARXIV.1412.6980

Kingma, D. P., & Welling, M. (2014). Auto-encoding variational bayes. In *2nd international conference on learning representations, ICLR 2014—Conference track proceedings*.

Koch, K., Wagner, G., Dahnke, R., Schachtzabel, C., Schultz, C., Roebel, M., Güllmar, D., Reichenbach, J. R., Sauer, H., & Schlösser, R. G. M. (2010). Disrupted white matter integrity of corticopontine-cerebellar circuitry in schizophrenia. *European Archives of Psychiatry and Clinical Neuroscience*, *260*(5), 419–426.

Kuhn, H. W. (1955). The Hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, *2*(1–2), 83–97.

Kuhn, H. W. (1956). Variants of the Hungarian method for assignment problems. *Naval Research Logistics Quarterly*, *3*(4), 253–258.

Lee, J. H., Lee, T. W., Jolesz, F. A., & Yoo, S. S. (2008). Independent vector analysis (IVA): Multivariate approach for fMRI group study. *NeuroImage*, *40*(1), 86–109.

Lee, M., & Pavlovic, V. (2021). Private-shared disentangled multimodal VAE for learning of latent representations. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 1692–1700.

Liang, M., Zhou, Y., Jiang, T., Liu, Z., Tian, L., Liu, H., & Hao, Y. (2006). Widespread functional disconnectivity in schizophrenia with resting-state functional magnetic resonance imaging. *Neuroreport*, *17*(2), 209–213.

Liu, S., Cai, W., Liu, S., Zhang, F., Fulham, M., Feng, D., Pujol, S., & Kikinis, R. (2015). Multimodal neuroimaging computing: A review of the applications in neuropsychiatric disorders. *Brain Informatics*, *2*(3), 167–180.

Ma, G., Lu, C. T., He, L., Philip, S. Y., & Ragin, A. B. (2017). Multi-view graph embedding with hub detection for brain network analysis. In *In: 2017 IEEE international conference on data mining (ICDM)* (pp. 967–972). IEEE.

Ma, S., Calhoun, V. D., Phlypo, R., & Adal, T. (2014). Dynamic changes of spatial functional network connectivity in healthy individuals and schizophrenia patients using independent vector analysis. *NeuroImage*, *90*, 196–206.

Peide, L., Sofuoglu, S. E., Maiti, T., & Aviyente, S. (2022). Coupled support tensor machine classification for multimodal neuroimaging data. arXiv preprint arXiv:220107683.

Plis, S. M., Amin, M. F., Chekroud, A., Hjelm, D., Damaraju, E., Lee, H. J., Bustillo, J. R., Cho, K. H., Pearlson, G. D., & Calhoun, V. D. (2018). Reading the (functional) writing on the (structural) wall: Multimodal fusion of brain structure and function via a deep neural network based translation approach reveals novel impairments in schizophrenia. *NeuroImage*, *181*, 734–747.

Qi, S., Calhoun, V. D., van Erp, T., Bustillo, J., Damaraju, E., Turner, J. A., du, Y., Yang, J., Chen, J., Yu, Q., Mathalon, D. H., Ford, J. M., Voyvodic, J., Mueller, B. A., Belger, A., McEwen, S., Potkin, S. G., Preda, A., Jiang, T., & Sui, J. (2018). Multimodal fusion with reference: Searching for joint neuromarkers of working memory deficits in schizophrenia. *IEEE Transactions on Medical Imaging*, *37*(1), 93–105.

Qi, S., Sui, J., Chen, J., Liu, J., Jiang, R., Silva, R., Iraji, A., Damaraju, E., Salman, M., Lin, D., Fu, Z., Zhi, D., Turner, J. A., Bustillo, J., Ford, J. M., Mathalon, D. H., Voyvodic, J., McEwen, S., Preda, A., … Calhoun, V. D. (2019). Parallel group ICA+ ICA: Joint estimation of linked functional network variability and structural covariation with application to schizophrenia. *Human Brain Mapping*, *40*(13), 3795–3809.

Sanislow, C. A., Ferrante, M., Pacheco, J., Rudorfer, M. V., & Morris, S. E. (2019). Advancing translational research using NIMH research domain criteria and computational methods. *Neuron*, *101*(5), 779–782.

Seitz, J., Rathi, Y., Lyall, A., Pasternak, O., del Re, E. C., Niznikiewicz, M., Nestor, P., Seidman, L. J., Petryshen, T. L., Mesholam-Gately, R. I., Wojcik, J., McCarley, R. W., Shenton, M. E., Koerte, I. K., & Kubicki, M. (2018). Alteration of gray matter microstructure in schizophrenia. *Brain Imaging and Behavior*, *12*(1), 54–63.

Shergill, S. S., Kanaan, R. A., Chitnis, X. A., O'Daly, O., M.Sc., Jones, D. K., Frangou, S., Williams, S. C. R., Ph.D., Howard, R. J., M.R.C.Psyc, Barker, G. J., Ph.D., Murray, R. M., F.R.C.Psyc, & McGuire, P., F.R.C. Psyc. (2007). A diffusion tensor imaging study of fasciculi in schizophrenia. *American Journal of Psychiatry*, *164*(3), 467–473.

Shi, Y., Paige, B., & Torr, P. (2019). Variational mixture-of-experts autoencoders for multi-modal deep generative models. *Advances in Neural Information Processing Systems*, *32*.

Silva, R. F., Plis, S. M., Adal, T., Pattichis, M. S., & Calhoun, V. D. (2020). Multidataset independent subspace analysis with application to multimodal fusion. *IEEE Transactions on Image Processing*, *30*, 588–602.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, *15*(1), 1929–1958.

Sui, J., He, H., Pearlson, G. D., Adali, T., Kiehl, K. A., Yu, Q., Clark, V. P., Castro, E., White, T., Mueller, B. A., Ho, B. C., Andreasen, N. C., & Calhoun, V. D. (2013). Three-way (N-way) fusion of brain imaging data based on mCCA+ jICA and its application to discriminating schizophrenia. *NeuroImage*, *66*, 119–132.

Sui, J., Jiang, R., Bustillo, J., & Calhoun, V. (2020). Neuroimaging-based individualized prediction of cognition and behavior for mental disorders and health: Methods and promises. *Biological Psychiatry*, *88*(11), 818–828.

Upthegrove, R., Marwaha, S., & Birchwood, M. (2017). Depression and schizophrenia: Cause, consequence, or trans-diagnostic issue? *Schizophrenia Bulletin*, *43*(2), 240–244.

van der Maaten, L., & Hintom, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research*, *9*(11), 2579–2605.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., … Vázquez-Baeza, Y. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in python. *Nature Methods*, *17*(3), 261–272.

Walterfang, M., Wood, S. J., Velakoulis, D., Copolov, D., & Pantelis, C. (2005). Diseases of white matter and schizophrenia-like psychosis. *Australian & New Zealand Journal of Psychiatry*, *39*(9), 746–756.

Wu, L., Calhoun, V. D., Jung, R. E., & Caprihan, A. (2015). Connectivity-based whole brain dual parcellation by group ICA reveals tract structures and decreased connectivity in schizophrenia. *Human Brain Mapping*, *36*(11), 4681–4701.

Wu, Y., & He, K. (2018). Group normalization. Proceedings of the European conference on computer vision (ECCV), pp. 3–19.

Yu, Q., Plis, S. M., Erhardt, E. B., Allen, E. A., Sui, J., Kiehl, K. A., Pearlson, G., & Calhoun, V. D. (2012). Modular organization of functional network connectivity in healthy controls and patients with schizophrenia during the resting state. *Frontiers in Systems Neuroscience*, *5*, 103.

Zhang, W., Zhan, L., Thompson, P., & Wang, Y. (2020). *Deep representation learning for multimodal brain networks* (pp. 613–624). International Conference on Medical Image Computing and Computer-Assisted Intervention Springer.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Geenjaar, E. P.T., Lewis, N. L., Fedorov, A., Wu, L., Ford, J. M., Preda, A., Plis, S. M., & Calhoun, V. D. (2023). Chromatic fusion: Generative multimodal neuroimaging data fusion provides multi-informed insights into schizophrenia. *Human Brain Mapping*, *44*(17), 5828–5845. https://doi.org/10.1002/hbm.26479