



# Dynamic Interrogation of Stochastic Transcriptome Trajectories Using Disease Associated Genes Reveals Distinct Origins of Neurological and Psychiatric Disorders

Theodoros Bermperidis<sup>1</sup>, Simon Schafer<sup>2</sup>, Fred H. Gage<sup>2</sup>, Terrence Sejnowski<sup>3</sup> and Elizabeth B. Torres<sup>1,4,5\*</sup>

<sup>1</sup> Sensory Motor Integration Laboratory, Department of Psychology, Rutgers University, Piscataway, NJ, United States, <sup>2</sup> Genetics Laboratory, The Salk Institute for Biological Studies, La Jolla, CA, United States, <sup>3</sup> Computational Neurobiology Laboratory, The Salk Institute for Biological Studies, La Jolla, CA, United States, <sup>4</sup> Computational Biomedicine Imaging and Modeling Center, Department of Computer Science, Rutgers University, Piscataway, NJ, United States, <sup>5</sup> Rutgers Center for Cognitive Science, Rutgers University, Piscataway, NJ, United States

## OPEN ACCESS

### Edited by:

Ming Li,  
Kunming Institute of Zoology (CAS),  
China

### Reviewed by:

Carlos P. Fitzsimons,  
University of Amsterdam, Netherlands  
Dennis Grayson,  
University of Illinois at Chicago,  
United States

### \*Correspondence:

Elizabeth B. Torres  
ebtorres@psych.rutgers.edu

### Specialty section:

This article was submitted to  
Neurogenomics,  
a section of the journal  
Frontiers in Neuroscience

Received: 26 February 2022

Accepted: 22 April 2022

Published: 02 June 2022

### Citation:

Bermperidis T, Schafer S,  
Gage FH, Sejnowski T and Torres EB  
(2022) Dynamic Interrogation  
of Stochastic Transcriptome  
Trajectories Using Disease Associated  
Genes Reveals Distinct Origins  
of Neurological and Psychiatric  
Disorders.  
Front. Neurosci. 16:884707.  
doi: 10.3389/fnins.2022.884707

The advent of open access to genomic data offers new opportunities to revisit old clinical debates while approaching them from a different angle. We examine anew the question of whether psychiatric and neurological disorders are different from each other by assessing the pool of genes associated with disorders that are understood as psychiatric or as neurological. We do so in the context of transcriptome data tracked as human embryonic stem cells differentiate and become neurons. Building upon probabilistic layers of increasing complexity, we describe the dynamics and stochastic trajectories of the full transcriptome and the embedded genes associated with psychiatric and/or neurological disorders. From marginal distributions of a gene's expression across hundreds of cells, to joint interactions taken globally to determine degree of pairwise dependency, to networks derived from probabilistic graphs along maximal spanning trees, we have discovered two fundamentally different classes of genes underlying these disorders and differentiating them. One class of genes boasts higher variability in expression and lower dependencies (High Expression Variability-HEV genes); the other has lower variability and higher dependencies (Low Expression Variability-LEV genes). They give rise to different network architectures and different transitional states. HEV genes have large hubs and a fragile topology, whereas LEV genes show more distributed code during the maturation toward neuronal state. LEV genes boost differentiation between psychiatric and neurological disorders also at the level of tissue across the brain, spinal cord, and glands. These genes, with their low variability and asynchronous ON/OFF states that have been treated as gross data and excluded from traditional analyses, are helping us settle this old argument at more than one level of inquiry.

**Keywords:** embryonic stem cells, transcriptome, neurological, psychiatric, tissues, autism, Parkinson, schizophrenia

## INTRODUCTION

The question of whether a distinction should exist between psychiatric and neurological disorders predates the time when psychiatry was not even a formal discipline as we know it today. Back then, motor movements were used as criteria to identify mental disorders, by observing and describing patients in a motor-informed way (e.g., catatonic, hyperactive, etc.) (Rogers, 1992). Under the spell of Freud's psychoanalyses and following Descartes's dualism, this type of physical-motor criterion lost influence in favor of elaborate descriptions of mental and emotional states inferred by other, non-motor-based criteria. There was more judgment added on to the perception of the patient; for example, terms such as deviant, opposing, defiant, socially inappropriate, behaviors, etc., entered the descriptions of children with atypical neurodevelopment. This judgment took place solely based on external observation, without any additional description of internal states of their nervous systems.

The distinction broadened between mental illness and disorders that affected the person's function beyond dysfunction of the central nervous system, thus prolonging the ongoing physiological and medical debates (Mehta, 2011); it also impacted the perception of other members of society with regards to one or the other (Torres, 2018). A recent revival of this debate underscores the side of the argument that psychiatric disorders are not just "mental" but are physical, too, identifying neurobiological substrates of mental illness (Goodkind et al., 2015). These substrates are in line with the current US National Institute of Mental Health Research Domain Criteria (RDoC), a framework that cuts across research domains (Bernard and Mittal, 2015) but still has room for improvement (Huys et al., 2016; Torres et al., 2016; Friston et al., 2017; Torres, 2020, 2021).

An example of the brain's affected tissues that are amenable to separating psychiatric from neurological disorders is provided in **Figure 1A** and results from this recent resurgence of the debate (Crossley et al., 2015). Yet, others have contested such neuroimaging distinctions between the disorders, on the grounds that medications can alter brain structures (David and Nicholson, 2015). Specifically, the argument is centered on the ambiguity that medication brings to the studies that are based on brain structure by, for example, increasing basal ganglia volume or increasing volume loss in general, in the case of traditional antipsychotics (David and Nicholson, 2015).

The interactions between diagnosis and medication are also mentioned in the Diagnostic Statistical Manual DSM-5 to justify the exclusion of motor criteria from diagnosis. Nevertheless, several of the mental disorders on a spectrum, like autism spectrum disorder (ASD), attention deficit hyperactivity disorder (ADHD), and schizophrenia do have functional neuromotor issues with neurobiological bases, i.e., of the neurological type, even when medication was never used (Torres and Denisova, 2016) or was sparsely used (Nguyen et al., 2016). Thus, the confounds between medication and psychiatry- or neurology-based diagnoses are palpable at the clinical level and confusing at the level of basic brain research.

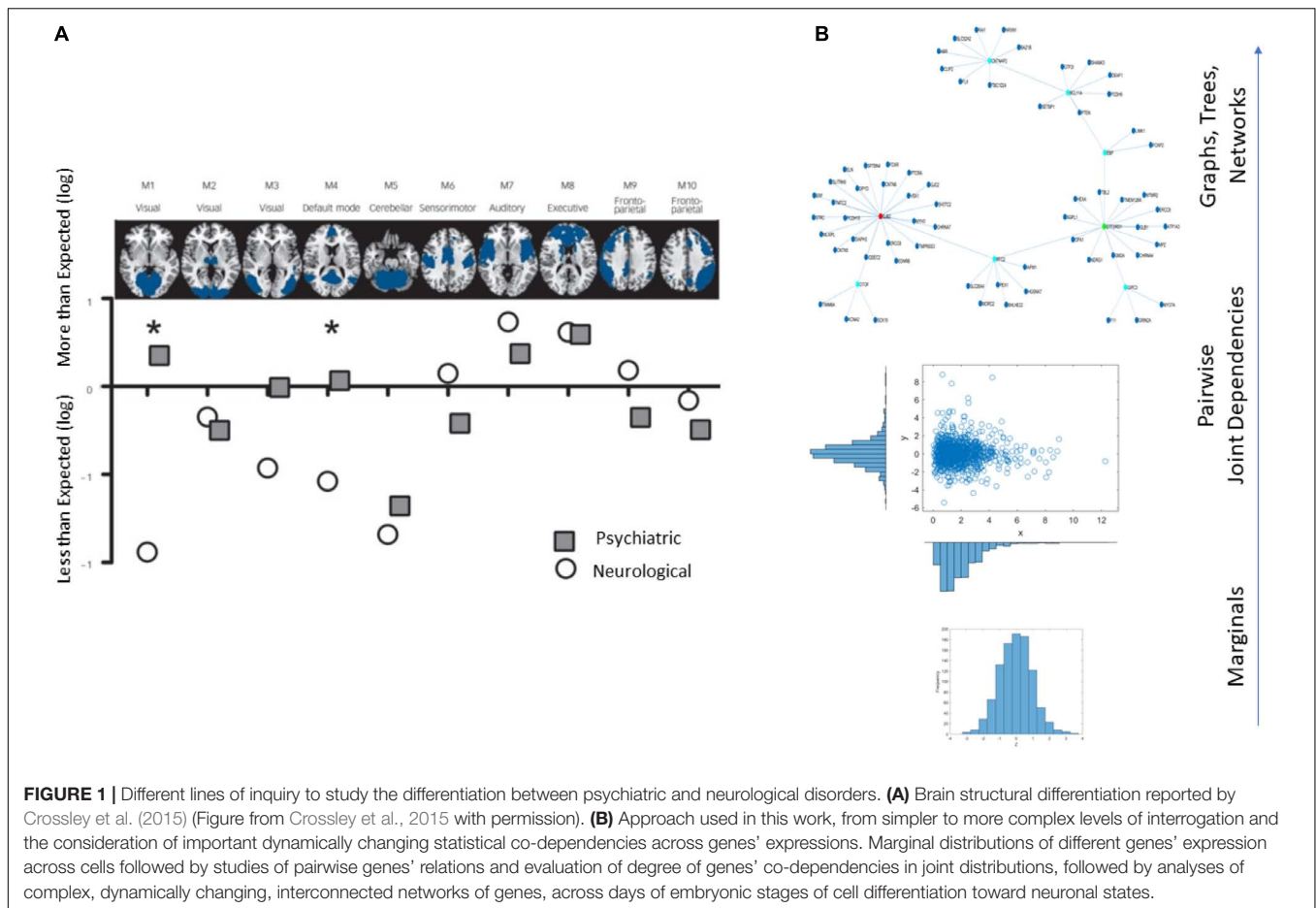
One avenue that we could explore to try and distinguish psychiatric and neurological disorders is by re-examining brain (and bodily tissues) from the standpoint of dynamically changing

gene expression in early embryonic stages of pluripotency, as cells transition into neuronal classes. In this context, we could use different levels of inquiry. For example, we could interrogate the genes with an eye on their roles in fundamental processes at the molecular or channel level, or perhaps at the systems level or at the level of tissues, etc., not as a role of the gene in isolation but rather as its role with respect to interactions with other genes. Regardless of the framework of choice, addressing possible differentiation between psychiatric and neurological disorders through genes' dynamic interactions and their expressions on brain and bodily tissues critical for the person's functioning may have new utility to aid in developing targeted treatments. Such treatments may be precisely aimed at mitigating such adverse effects on the brain and on the control of the movements that make up the behaviors examined by these observational diagnoses in the first place.

In this paper we leverage recent advances in the modeling of neurodevelopmental stages involving human embryonic stem cells (hESC), which have made transcriptome data from early development available to the scientific community. Such sharing of data from cultures validated by primary developing tissue offers new opportunities to advance analytical and visualization tools that can potentially facilitate the study of the dynamics of cell differentiation across multiple developmental stages. It can help us shed light on the question of differentiating pools of genes associated with disorders of the nervous systems that may or may not rise to the level of mental illness.

An example of such open access work is by Yao et al. (2017), which modeled the early stages of human brain development, including early regional patterning and lineage specification. These authors described cell characterizations amenable to providing benchmark datasets to advance our understanding of the origins of disease of the human brain. Here we use their data to design new visualization tools inclusive of all genes' fates in the transcriptome and genes' states across differentiation of self-emerging patterns recorded several times over 54 days. The results available from single-cell RNA sequencing (scRNA-seq) or single cell transcriptomics offer gene-expression data from tens of thousands of genes across hundreds of cells evolving and differentiating toward neuronal stages. These data, combined with identification of disease-associated genes and their expression on human brain and bodily tissues, may help us track the origins of differentiation between psychiatric and neurological disorders.

Analyses of such data often entails dimensionality reduction and visualization of the reduced set [e.g., after Principal Component Analyses, PCA initialization and *t*-distributed stochastic neighbor embedding, *t*-SNE (van der Maaten and Hinton, 2008)]. Often, during several of the steps leading to the embedding and visualization in much lower dimensional spaces (of two or three dimensions), thousands of genes may be discarded owing to low variability and/or asynchronous expression across various reading days. These data that are discarded may, however, be key to cases where atypical development takes place. Consequently, genes that may be critical to early development toward neuronal stages could be potentially disregarded by current popular methods like *t*-SNE. This approach would miss an opportunity to examine



the transcriptome data from the vantage point of inter-related nodes in a network, using a stochastic approach that goes beyond locally selected neighboring interactions of genes with systematic variability to leverage (and understand) the dynamic, asynchronous nature of many otherwise discarded genes during pluripotent neuronal differentiation.

We propose new methodologies (Figure 1B) that treat a set of genes as a network entity whose parts interact with each other over the course of cell development. To that end, we use a layered approach. First, we identified genes associated with a plurality of psychiatric and neurological disorders, and which also overlapped, thus being associated with phenotypes that are considered comorbid with, for example, autism spectrum disorders (ASD), attention deficit hyperactivity disorder (ADHD), cerebral palsy (CP), etc. (Torres, 2020, 2021). Then, we consider the cumulative expression of the genes across four readings through 54 days, as the cells transition to neuronal classes. For each gene, we derived marginal distributions of expression across cells and tracked pairwise dependencies to interrogate the full transcriptome dynamically on each reading day. We did so within another layer of inquiry, as the genes formed part of a probabilistic interconnected graph.

This network of interacting interconnected genes associated with a plurality of psychiatric and neurological disorders cannot be separated into a disjointed collection of data points, since the

topological properties of the graph determine expression and differentiation. We therefore found relationships between local and global properties of this network by defining metrics that quantified the “fate” of each gene during cell differentiation and the degree of interdependency between all cells. Furthermore, our approach was dynamic, i.e., we looked at gene expressions on multiple days for a culture of cells. This approach allowed us to shed some light on how changes in the “state” of the expressions of different genes during the embryonic stages determined the clinical phenotype and characterized different pathologies of the CNS.

This characterization based on the fate and state of the genes' expressions led to the proposition of a general mechanism and new paradigm that traces the origins of differentiation and commonalities between neurological and psychiatric disorders back to the early stages of embryonic development. This is the point when cells differentiate and become fully matured neurons that will make up the different systems of the nervous system. In this sense, we transformed the current psychiatric vs. neurological disorders debate into an opportunity to explore when, during these early embryonic stages, the genes expressed as one disorder or another as a function of their degree of interdependencies. We discuss the implications of our results while considering the notion that gross data such as low-variability and asynchronous genes expression, which

are often discarded as superfluous, may in fact hold the key to many unknown aspects of neurological and/or psychiatric disorders. Developing new methods to harvest and utilize their dynamically changing stochastic activities may be critical to understanding the mechanisms guiding us in the design of treatments to cure diseases.

## DATASETS AND CODES

Single-cell RNA-seq data during neural differentiation of hESCs were provided by the study of Yao et al. (2017). They revealed a multitude of neural cell classes with a range of early brain regional identities. They analyzed 2,684 cells with >20,000 transcripts, as assessed by unique molecular identifiers (UMIs). Cell types were named by point of origin as progenitor (P), transitional (T), neuronal (N), or other tissue (O). We focussed on the evolution of 24,046 genes' expression across the neuronal type examined at days 12, 19, 40, and 54.

The code that implemented the Kernel Statistical Test of Independence and was used for the present analysis is available in Gretton and Gyorfi (2010) as free source material on their website (<https://www.gatsby.ucl.ac.uk/~gretton/>).

## MATERIALS AND METHODS

### Disorders

From the genetic database of the Simons Foundation Research Initiative (SFARI) we gathered the set of genes that, according to the literature, are linked to the behavioral diagnosis of Autism. We also compiled several sources found in the DisGeNet portal and identified the lists of genes linked to a variety of neurological [Fragile X-Associated Tremor/Ataxia Syndrome (FXTAS), Dystonia, Cerebral Palsy (CP), Ataxia Syndrome, Tourette's, and Late and Early Onset Parkinson's disease (White et al., 2012)] and psychiatric disorders [Schizophrenia, Depression, Obsessive Compulsive disorder (OCD), bipolar disorder, and Post-Traumatic Stress Disorder (PTSD)]. The psychiatric disorders under consideration are defined in the DSM, whereas the neurological ones are comorbid with autism spectrum disorders (ASD) across the human life span. Some disorders like CP and Tourette's are also included in the DSM.

### Transcriptome Data

Data matrices expressed as  $M$  cells  $\times$   $N$  genes containing at each entry transcription in log Counts/million. The matrix is transposed to express each of the 24,046 genes as it expresses across cells. This is represented as  $X_{N \times M}^D$ , where  $N$  is the number of genes and  $M$  the number of cells for day  $D$ . **Figure 2A** shows the counts across neurons at day 40 (top) and at day 54 (bottom) for one gene. Notice that cells may not necessarily be the same on each day, but we work on the genes' expression space. Next, we took the expression of each gene at each cell as peaks in a series (shown in red) and normalized them using Eq. 1. Notice that the cell order is preserved from the original matrix of cells by genes. It is not a temporal order (it is not

a time series), and as such, order does not matter in our calculations:

$$Normalized\ Count = \frac{count_i}{count_i + \frac{Avrg_{Global}Count}{Max_{Global}Count}} \quad (1)$$

here  $count_i$  is the count value of the  $gene_i$  and  $Avrg_{Global}Count$  is the overall average of the matrix of values taken along the columns and the rows.  $Max_{Global}Count$  is the maximum count value, also taken globally across the matrix values. We take each such value and apply Eq. 1 to scale all expressions of that gene across all cells, each day. The output of the normalization is shown in **Figure 2B** for cells with expression close to 1. These are cells where the ratio of  $Avrg_{Global}Count$  to  $Max_{Global}Count$  is very small, so the denominator is only a small margin greater than the numerator. The inset stem plot in **Figure 2C** (inside the top and bottom histograms) represents all values of the gene across all the cells, scaled as spikes ranging between 0 and 1. These include values for which the ratio of  $Avrg_{Global}Count$  to  $Max_{Global}Count$  is large and the overall normalized value is small. There is not a temporal order, it is just a series of fluctuations whereby the frequency histogram of all those values (smaller and larger) is of interest.

We then obtained the similarity distance between the two frequency histograms corresponding to the gene (across all cells read out each day) using the Earth Mover's Distance (EMD) (Monge, 1781; Rubner et al., 1998).

### Genes' Fate: Quantifying It Through the Cumulative Earth Mover's Distance

The EMD is a measure of the distance between two probability distributions and is the 1st Wasserstein distance from the family of Wasserstein distance metrics. Essentially, the EMD is the minimal cost of transforming one distribution into the other.

Consider the unknown probability spaces  $([0, +\infty], F_i^{D_A}, P_i^{D_A})$  and  $([0, +\infty], F_i^{D_B}, P_i^{D_B})$  that define the statistical behaviors and corresponding probability distributions for a gene on two different days,  $D_A$ , and  $D_B$  during cell differentiation. The probability distributions can be approximated through histogram fitting from the available normalized cell expression data. Then,  $EMD_{i,A \rightarrow B}$  is the EMD between  $P_i^{D_A}$  and  $P_i^{D_B}$  and is indicative of the departure of the statistical behavior of the gene  $i$  as we move from day  $D_A$  to  $D_B$ . Consider the days 12, 19, 40, and 54 of hESC differentiation. Then we define the quantity:

$$CumEMD_i = EMD_{i,12 \rightarrow 19} + EMD_{i,19 \rightarrow 40} + EMD_{i,40 \rightarrow 54} \quad (2)$$

as the cumulative EMD distance of the gene or the "fate" of the role of that gene throughout embryonic stem cell development and differentiation. Then, the average EMD or "fate" of a set of genes  $i=1, \dots, N$  associated with a particular pathology is the quantity:

$$AverageCumEMD = \frac{\sum_{i=1}^N CumEMD_i}{N} \quad (3)$$

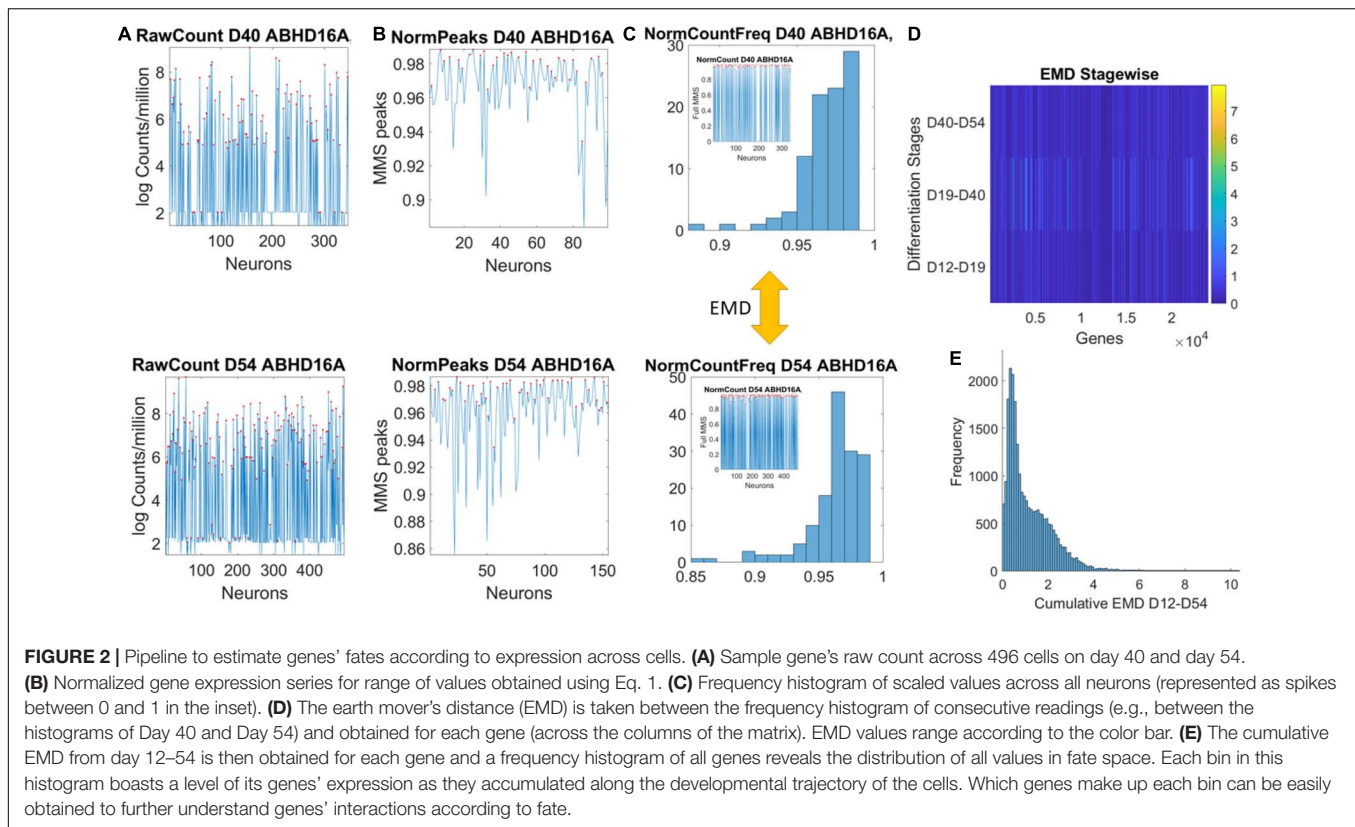


Figure 2D shows a color matrix with all genes across the columns (horizontal axis) and three rows representing, for each gene, the EMD obtained between two consecutive days.

We focused here on graphical modeling techniques and kernel-driven statistical independence hypothesis testing to determine the level at which these genes co-depend (Gretton and Györfi, 2010). This kernel technique was used to build a parameter space representing 16 disorders sorted out according to their levels of statistical co-dependency at the earliest and latest stages of the cell evolution. Our data of interest were the cumulative EMD traveled representing the gene's fate.

### Kernel Statistical Test to Determine the Level of Interdependence Between Genes Associated With Each Disorder at Different Times During Cell Maturation

On a specific day during cell maturation, for any two genes A and B we had available their expression levels in N cells. We wanted to test whether the expression of gene A was statistically independent from the expression of gene B. Before we present the solution to this problem, let us first introduce the general framework for measuring independence, based on cross-covariance operators in *Reproducing Kernel Hilbert Spaces* (RKHSs).

#### Cross-Covariance Operator and Hilbert–Schmidt Independence Criterion

Different methods for measuring statistical independence between random vectors have been proposed over the past

decades. Non-parametric approaches to this problem can be traced back to Hoeffding (1948), when he proposed a test statistic that depended on the rank order of the identically and independently distributed sample data. Techniques that constructed statistics based on empirical characteristic functions were later developed. Modern methods introduced the concept of Kernel Independence Measures, which have found applications in Independent Component Analysis (Bell and Sejnowski, 1996, 1997), fitting graphical models and feature selection. However, such measures do not necessarily ensure statistical significance. Hence, we decided to use a Kernel Statistical Test of Independence that was developed by Gretton and Györfi (2010), which allowed us to perform hypothesis testing on whether two datasets were independent, and which we applied on the gene expression dependence problem. We briefly present their methodology in the next few paragraphs.

Consider the Hilbert space  $F$  of functions from a measurable space  $X \rightarrow \mathbb{R}$ . To each point  $x \in X$  there corresponds an element  $\varphi(x) \in F$  such that  $\langle \varphi(x), \varphi(x') \rangle_F = k(x, x')$  where  $k: X \times X \rightarrow \mathbb{R}$  is a positive definite kernel. If we assume that  $F$  is separable, then  $F$  is a RKHS.

Note, that  $F$  is the completion of the set of all functions that are linear combinations of these feature functions. To evaluate the value of any function  $f \in F$  at some point  $x \in X$  one can simply take the inner product between the function  $f$  and the feature function  $\varphi(x)$  mapping of the point  $x$ . This is known as the Reproducing Property, hence the term Reproducing Kernel Hilbert Space. Similarly, we define a RKHS space  $G$  of functions

from a measurable space  $Y \rightarrow R$  with feature map  $\psi(\cdot)$  and kernel  $l(\cdot)$ .

Now, assume the probability spaces  $(X, \mathcal{F}_x, P_x)$  and  $(Y, \mathcal{F}_y, P_y)$  where  $\mathcal{F}_x, \mathcal{F}_y$  are the Borel  $\sigma$ -fields on  $X, Y$ , respectively, and  $P_x, P_y$  the corresponding probability measures. Then, for any two functions  $f \in F$  and  $g \in G$  the cross-covariance operator  $C_{xy} : G \rightarrow F$  is defined as:

$$\langle f, C_{xy}g \rangle = E \left( [f(x) - E(f(x))] [g(x) - E(g(x))] \right) \quad (4)$$

Or with respect to the feature mappings  $\varphi(x), \psi(y)$ :

$$C_{xy} = E_{xy} \left[ (\varphi(x) - E_x \varphi(x)) \otimes (\psi(y) - E_y \psi(y)) \right] \quad (5)$$

It can be shown that  $X$  and  $Y$  are independent if and only if the largest singular value of the operator  $C_{xy}$  is zero. As a measure of independence, the authors consider the Hilbert–Schmidt norm (i.e., the Hilbert–Schmidt Independent Criterion, HSIC) of  $C_{xy}$ , which is equal to the sum of squared singular values of  $C_{xy}$  and has a population expression (Gretton and Györfi, 2010):

$$HSIC(P_{xy}, F, G) = E_{xx'yy'} [k(x, x') l(y, y')] + E_{xx'} [k(x, x')] E_{yy'} [l(y, y')] - 2E_{xy} [E_{x'} [k(x, x')] E_{y'} [l(y, y')]] \quad (6)$$

where  $x'$  denotes an independent copy of  $x$  and  $k(\cdot)$  and  $l(\cdot)$  are the kernels previously defined. The authors derive an empirical estimate of this independence criterion that follows the V-statistics and has expression:

$$HSIC_b(Z) = \frac{1}{m^2} \text{trace}(KHLH) \quad (7)$$

where  $Z$  is a sample of  $(x, y)$  pairs drawn independently from the distribution of  $X \times Y$ , with size  $m$ ,  $K$  is the  $m \times m$  matrix with entries  $k_{ij}$ ,  $L$  is the  $m \times m$  matrix with entries  $l_{ij}$  and  $H = I - \frac{1}{m} \vec{1} \vec{1}^T$ , where  $\vec{1}$  is a row vector of ones.

Then, they proceed to construct a statistical hypothesis testing protocol to test whether  $X$  is independent of  $Y$  based on samples  $(x, y)^m$  drawn from the probability space  $(X \times Y, \mathcal{F}_x \times \mathcal{F}_y, P_{xy})$ . The null hypothesis is  $H_0 : P_{xy} = P_x P_y$  and the alternative hypothesis  $H_1 : P_{xy} \neq P_x P_y$ .

Finally, they approximate the independence criterion with a gamma distribution:

$$mHSIC(Z) \sim \frac{x^{k-1} e^{-\frac{x}{\theta}}}{k^k \Gamma(k)}, \quad (8)$$

$$\text{where } k = \frac{(E(HSIC_b(Z)))^2}{\text{var}(HSIC_b(Z))}, \quad \theta = \frac{m \text{var}(HSIC_b(Z))}{E(HSIC_b(Z))}$$

If  $mHSIC(Z)$  is above the threshold determined by the level of significance that we choose for the test, the null hypothesis is rejected. **Supplementary Figure 0** depicts this pipeline.

### The Construction of Gene Expression Statistical Dependency Networks for Days 12, 19, 40, and 54 of Neural Cell Maturation

For each disorder, we have a set of genes associated with it (extracted from DisGeNet). On a particular day  $D$  of cell

maturation, we have the data  $X_{N \times M}^D$ , where  $N$  is the number of genes and  $M$  is the number of cells. Each row of our data refers to a specific gene and each column to a specific cell.

Define the unknown probability spaces  $([0, +\infty), \mathcal{F}_i, P_i)$  and  $([0, +\infty), \mathcal{F}_j, P_j)$  for the expressions of any two genes  $i$  and  $j$ ,  $(i, j = 1, \dots, N)$  and the probability space  $([0, +\infty)^2, \mathcal{F}_i \times \mathcal{F}_j, P_{ij})$  for the joint expression of the two genes. Here,  $[0, +\infty)$  and  $[0, +\infty)^2$  are the measurable spaces for the gene expressions and joint gene expressions, respectively,  $\mathcal{F}_i, \mathcal{F}_j$  the Borel  $\sigma$ -fields generated by the measurable spaces of gene expressions of  $i$  and  $j$ ,  $P_i, P_j$  the corresponding probability measures for the two genes and  $P_{ij}$  the joint probability measure.

Consider the sample  $Z = (X_{i,*}^D, X_{j,*}^D)$ , i.e., the pairs of the two gene expressions in the cells. Choosing a level  $\alpha$  of statistical significance, we can apply the Kernel method on the sample  $Z$  and determine whether the genes  $i$  and  $j$  are independent of each other on that specific day.

We perform the independency test on all undirected pairs  $(i, j)$ ,  $i = 1 \dots N - 1, j = i + 1, \dots, N$  of genes and we construct the graph:

$G = (V, E)$ ,  $V$  is the set of nodes and  $E$  the set of edges,  $|V| = N$ , where  $|\cdot|$  denotes the cardinality of a set. An edge belongs to the graph, i.e.,  $e_{ij} \in E$  if and only if the genes  $i$  and  $j$  are statistically dependent according to the kernel independency test.

If the graph were fully connected it would have number of edges  $\frac{N(N-1)}{2}$ , and in this case all genes would be fully dependent upon one another. In search of a metric that shows how interconnected the genes related to the disorder of interest are, we define the *dependency index*:

$$DI = \frac{2|E|}{N(N-1)} \quad (9)$$

Therefore, for a specific disorder  $d$ , by constructing the dependency graphs  $G_{12}^d, G_{19}^d, G_{40}^d, G_{54}^d$  for each day of the disorder we can derive the dependency indexes  $DI_{12}^d, DI_{19}^d, DI_{40}^d, DI_{54}^d$ . Furthermore, we define:

$$\begin{aligned} \text{initial state} &= DI_{12}^d \\ \text{final state} &= DI_{54}^d \\ \text{Dependency Increase} &= DI_{54}^d - DI_{12}^d \end{aligned} \quad (10)$$

Then, we can map each disorder to a parameter space of dependencies throughout the course of the cell development to track how the interdependence between the genes associated with each disorder evolves through time, from the state of pluripotency to the state of full neural maturation. This allows us to stratify the spectrum of neurological and psychiatric disorders with regards to the complexity of their genotypical expressions.

Using the database DisGeNet, we identified the genes that are associated with each of the following disorders: Fragile X-Associated Tremor/Ataxia Syndrome (FXTAS), Dystonia, Cerebral Palsy, Ataxia Syndrome, Tourette’s, Late Onset Parkinson’s disease (Late PD) and Early Onset Parkinson’s disease (Early PD) and Schizophrenia, Depression, Obsessive Compulsive disorder (OCD), bipolar disorder, and Post-Traumatic Stress Disorder (PTSD).

For days 12, 19, 40, and 54 of the hESC differentiation to neural cells, we had the expressions of 24,046 genes from the human genome in log TCM (log transcripts count per million). For days 12, 19, 40, and 54, we had available the expression of all those genes in 263, 168, 346, and 495 different cells, respectively. Therefore, for the sets of genes associated with each of the disorders of interest, we had the datasets  $X_{N \times 263}^D$ ,  $X_{N \times 168}^D$ ,  $X_{N \times 346}^D$ , and  $X_{N \times 495}^D$  for the 4 days, where D denotes the disorder of interest and N the number of genes associated with it. From these datasets we calculated the Dependency Indexes (see Section “Materials and Methods”) for each day. **Figure 3A** shows the schematics of this procedure.

## Network Analyses

### Undirected Graphical Modeling for Evaluating the Dependency Between Gene Expressions

Graphical models have been extensively researched and used to describe statistical dependencies between random variables. These models can be either undirected graphs or directed graphs; in the latter case, we could derive cause-effect relationships between the variables. In the current project we were interested in undirected graphical models.

Formally, for any set of random variables  $X = (X_1, X_2, \dots, X_N)$ , a graphical model attempts to associate the joint random vector  $X$  drawn from the probability space  $(\Omega_1 \times \Omega_2 \times \dots \times \Omega_N, F_1 \times F_2 \times \dots \times F_N, P_X)$  with a graph  $G = (V, E)$ , where  $V$  stands for vertices and  $E$  for edges. Here,  $\Omega_1, \Omega_2, \dots, \Omega_N$  are the sample spaces for each random variable and  $\Omega_1 \times \Omega_2 \times \dots \times \Omega_N$  the joint space of the random vector  $X$ .  $F_1, F_2, \dots, F_N$  are the corresponding generated Borel  $\sigma$ -fields to denote the sets of all possible random events for each random variable and  $P_X$  is the probability measure on the random vector  $X$ . The set of nodes  $V$  represents the random variables  $X_1, X_2, \dots, X_N$  drawn from the probability spaces  $(\Omega_1, F_1, P_{X_1}), (\Omega_2, F_2, P_{X_2}), \dots, (\Omega_N, F_N, P_{X_N})$ , with their respective Borel  $\sigma$ -fields and probability measures. An edge  $e_{ij} \in E$  if and only if the random variables  $X_i$  and  $X_j$  depend on each other.

If two nodes  $u_i$  and  $u_j$  are not connected with an edge it implies that the two variables  $X_i$  and  $X_j$  are conditionally independent, i.e., statistically independent *given* all other nodes:

$$X_i \perp X_j \mid X_{V/\{u_i, u_j\}}$$

This property of the graphical model is known as the *global Markov property*.

### General Estimation of a Graphical Model Using Chow–Liu Trees

If we want to factorize the joint probability distribution in a dependency graph that has a tree structure, i.e., every two nodes are connected by no more than one path (in other words there are no loops in the graph), then the joint density of the random vector

$X$  factorizes with respect to the pair-wise joint and marginal densities as:

$$f(x) = \prod_{e_{ij} \in E} \frac{f_{ij}(X_i, X_j)}{f_i(X_i)f_j(X_j)} \prod_{u_i \in E} f_i(x_i) \tag{11}$$

It turns out that, in the case in which all variables are categorical and take values from a finite set, it is very easy to find the optimal tree that factorizes the joint distribution. Let  $N_x$  be the number of times a realization  $x$  of the random vector  $X$  appears in a collection of independent and identically distributed (*i.i.d.*) samples. The tree  $G$  that optimally factorizes  $X$ , given the sample data  $Z$  of size  $n$ , will be the one with the maximum log-likelihood (MLE):

$$G = \operatorname{argmax}_G L(G) = \sum_{x \in Z} N_x \log(\hat{f}_G(x)) \tag{12}$$

which turns out to be:

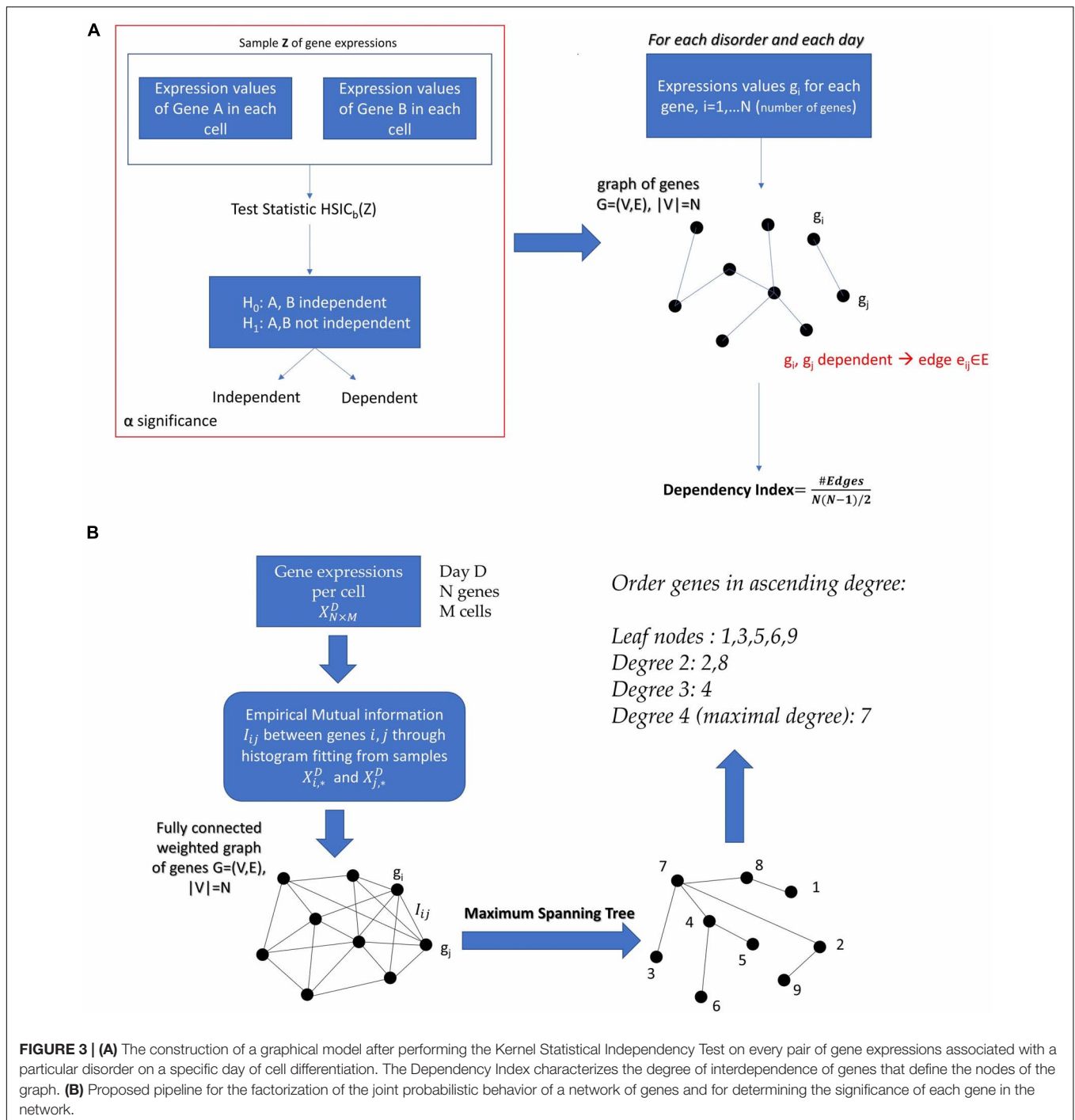
$$L(G) = n \sum_{e_{ij} \in E} I(\hat{f}_{ij}(X_i, X_j)) + C \tag{13}$$

where  $C$  is a constant and  $I(\hat{f}_{ij}(X_i, X_j))$  the empirical mutual information between  $X_i$  and  $X_j$ . Therefore, by choosing the appropriate subgraph  $G$  that maximizes the sum of the empirical mutual information estimates, we obtained the optimal tree structured graphical model. Since the model is a tree, we simply needed to find the Maximum Spanning Tree from the mutual information network, for example, by using Kruskal’s algorithm. The process we just described is known as the Chow–Liu algorithm and the extracted conditional dependency tree that factorizes  $X$  is the Chow–Liu tree.

### Gene’s Co-dependencies Graphical Models Spanning Chow–Liu Trees

For a specific set of diseases, in this case neurological and psychiatric, we had a set of genes associated with them. On a particular day  $D$  of cell maturation, we had the data  $X_{N \times M}^D$ , where  $N$  is the number of genes and  $M$  the number of cells. Each row of our data referred to a specific gene and each column to a specific cell. We treated each column (corresponding to a cell) as an *i.i.d.* sample drawn from the joint probability space of the expressions of that set of genes, and we wanted to generate the undirected graphical networks  $G_{12}, G_{19}, G_{40}$ , and  $G_{54}$  for days 12, 19, 40, and 54.

In the case of continuous variables, the methods used to estimate the Chow–Liu tree usually involves Kernel Density Estimation (KDE) of the joint and marginal probability densities (Gretton et al., 2007). However, in our case we wanted to factorize the joint probability density of a gene expression network with number of cells on the order of magnitude  $\sim 10^2$ . The application of KDE, given the dimensionality of the data (number of genes), would require in this case several samples far exceeding the available number of ESCs. Therefore, we resorted to extracting the Chow–Liu Trees by estimating the mutual information through binning and histograms on the available data (Drton and Maathuis, 2017). Once the Chow–Liu Tree corresponding to the



factorization of the gene set of interest was obtained, we ordered the nodes in ascending degree. The proposed methodology can be appreciated in **Figure 3B** of the Section “Materials and Methods.”

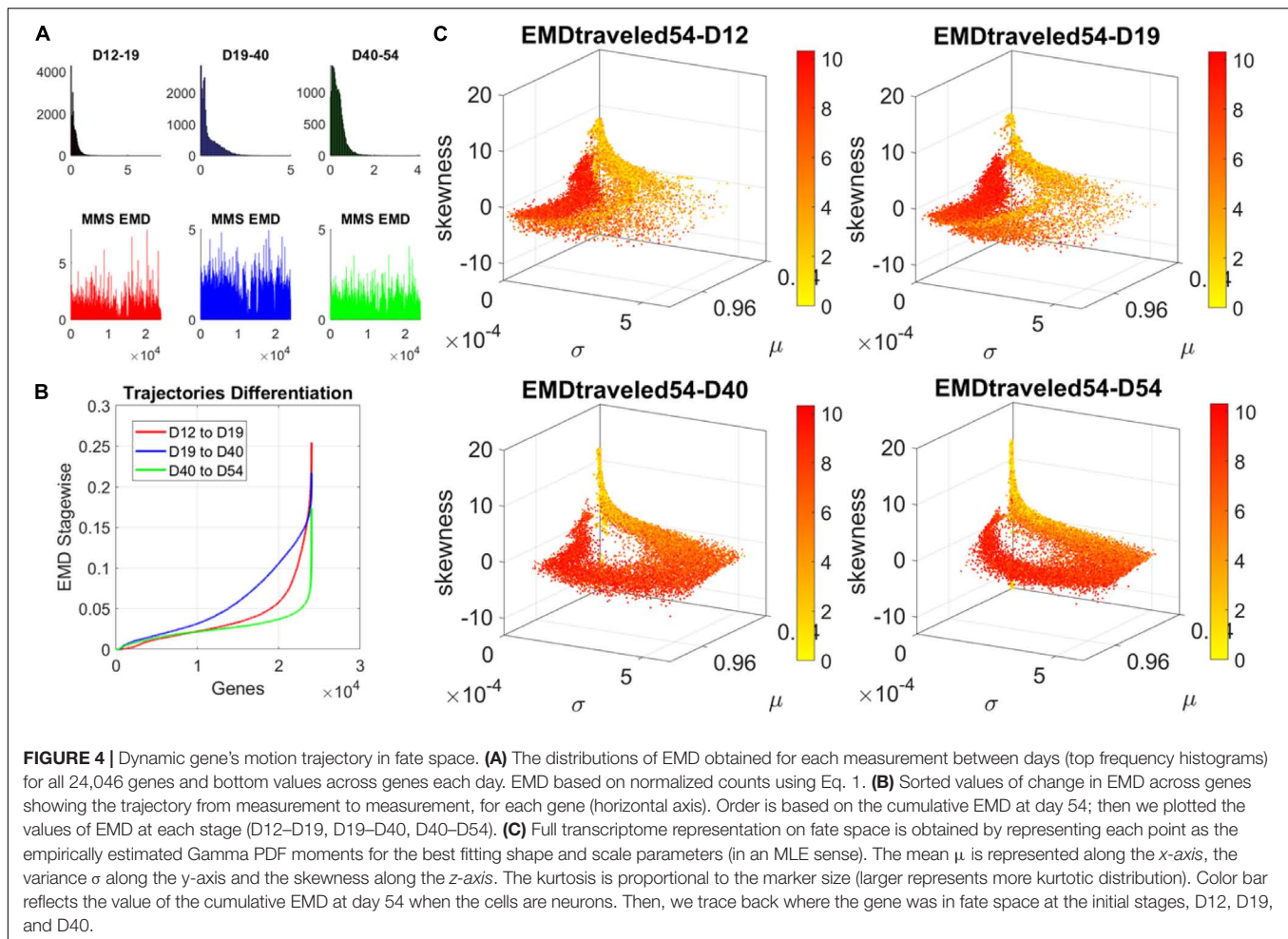
What do we achieve with this ordering? It is obvious that the higher the degree of a gene is, the more co-dependent its expression is with the expression of other genes in the network. This implies, in a statistical sense, that mutation or deletion of this gene is bound to immediately affect the expressions of many other genes. Note that this rationale simply states the co-dependency between the gene expressions, but the actual (causal) mechanisms

through which this statistical relationship takes place are open to investigation.

### The Gene’s State Through a Binary Code

We obtained the average degree across the network’s nodes each day. We then set it as the threshold value to determine ON or OFF state for the gene each day. Across 4 days, we had 16 possible binary states ( $2^4$ ) that provided the state trajectory of the gene (in addition to its fate). This information served to classify cells according to the genes type. To that end, for each cell, we obtained





the frequency histogram of counts corresponding to each class of genes, [0000], [0001], . . . , [1111]. Then, we used MLE to obtain the probability distribution best characterizing the histogram and obtained a parameter space to represent the shape and scale parameters thus determined as points along a scatter. Since each gene has a binary configuration, for a given day, we could then retake the cells  $\times$  genes matrix and ask which cells cluster in this space according to the 16 possible states.

## RESULTS

### Cumulative Earth Mover's Distance Captures Dynamic Evolution of the Genes' Expression in Fate Space

The contributions of each gene to the overall evolution of the hESCs as they reached neuronal stages were well captured by the stochastic characterization of their normalized count taken for each gene across cells each day. These marginal distributions can be seen evolving across all genes in **Figure 4A**, which shows the frequency histogram of all EMD values obtained for each comparison. **Figure 4B** shows the values sorted out across the

genes, providing a sense of the overall trajectory of changes in gene expression across all cells. Also notice that, since each day the number of cells changes, we normalized the EMD quantity to range between 0 and 1 when superimposing the data across days (**Figure 4B**).

We then accumulated the EMD value by summing up to day 54 and building a colormap to visualize the change across all genes in the transcriptome. For each gene, we took the frequency histogram of the normalized gene expression across the cells and, using maximum likelihood estimation (MLE), we obtained the continuous family of probability distributions that best fit the histogram in an MLE sense. This was the Gamma family, which spans a broad range of shape ( $a$ ) and scale ( $b$ ) values (ranging from the memoryless exponential  $a = 1$ , to distributions with heavy tails, to symmetric, Gaussian-like distributions). We estimated the Gamma moments of the empirical distribution corresponding to each gene. Each day we plotted them on a parameter space, whereby the mean is represented along the  $x$ -axis, the variance is represented along the  $y$ -axis, the skewness is represented along the  $z$ -axis, and the kurtosis is proportional to the size of the marker (higher kurtosis represented by larger marker size). We then colored each gene with the cumulative EMD at day 54, when the cells had reached neuronal state. This

enabled us to visualize the “motion” dynamics of the 24,046 genes in the transcriptome and, retrospectively, see where the most active and least active genes were located. This information is shown in **Figure 4C**, as the cells evolved to neuronal stages.

## Different Evolution of the Dependency Index Values Throughout Human Embryonic Stem Cells Maturation for Psychiatric vs. Neurological Disorders

We obtained the quantity Dependency Increase, as explained in the methods, by taking the shift in the Dependency Index at Day 12 versus the increase in the Dependency Index until full neuronal maturation at day 54. We then plotted that value for each disorder along the *y-axis* of a parameter space, as a function of the Dependency Index at the initial stage and at the final stage. This is represented as a vector field (rate of change) in **Figure 5A**. A pattern emerged across disorders whereby all disorders that were classified as neurological, except for Late Onset Parkinson’s Disease, tended to be characterized by a lower dependency during the first stage of cell differentiation and a high increase in dependency at the neuronal stage. The genes associated with these neurological disorders tended to increase their co-dependencies as the cells matured into neurons, but early in the process they were less co-dependent (lower values of Dependency Index on the *x-axis*). These results can be appreciated in **Figure 5B**.

In contrast, disorders classified as psychiatric in the DSM-5, as well as Late Onset Parkinson’s Disease (and except for Infantile Schizophrenia), tended to be characterized by a higher dependency during the first stage (higher values along the *x-axis*) but a lower increase in dependency as the cells matured into neurons. Tourette’s, a psychiatric disorder according to the DSM-5, was also traditionally classified as neurological and often labeled as ASD. In this parameter space, ASD seemed to lie on the border between neurological and neuropsychiatric disorders, whereas infantile schizophrenia (which used to be defined as autism earlier in DSM history) lined up with ASD along the early value of the Dependency Index, but with a much higher dependency increase as cells matured into neurons. We highlight this duality with the black edge of the marker in CP and Tourette’s on panel 5B using a parameter space that we explain next.

## Negative Correlation Trend Characterizes Genes Associated With Neurological and Psychiatric Disorders, With Complementary Features in Autism Spectrum Disorder and Parkinson’s Disease Associated Genes

For the genes associated with each disorder, we calculated the average cumulative EMD at Day 54 (*y-axis*) and plotted it as a function of the Dependency Index at Day 54 (*x-axis*) in **Figure 5B**. We noticed a negative trend whereby the variability of expression, as quantified by the cumulative EMD from measurement to measurement, tended to decrease for

Depression, PTSD, OCD, CP, Tourette’s (all DSM disorders). In this cluster, early and late PD, which are neurological disorders, appeared amid psychiatric DSM-classified disorders.

Neurological disorders such as FXTAS, FX, Dystonia, and Ataxia were high in cumulative EMD, thus implying higher cumulative changes in variability of genes’ expression. However, these genes tended to have lower dependency indexes than DSM-classified disorders. ASD, currently classified as a DSM psychiatric disorder, appeared among the neurological cluster with a high cumulative EMD at day 54, but a lower dependency index during this final state.

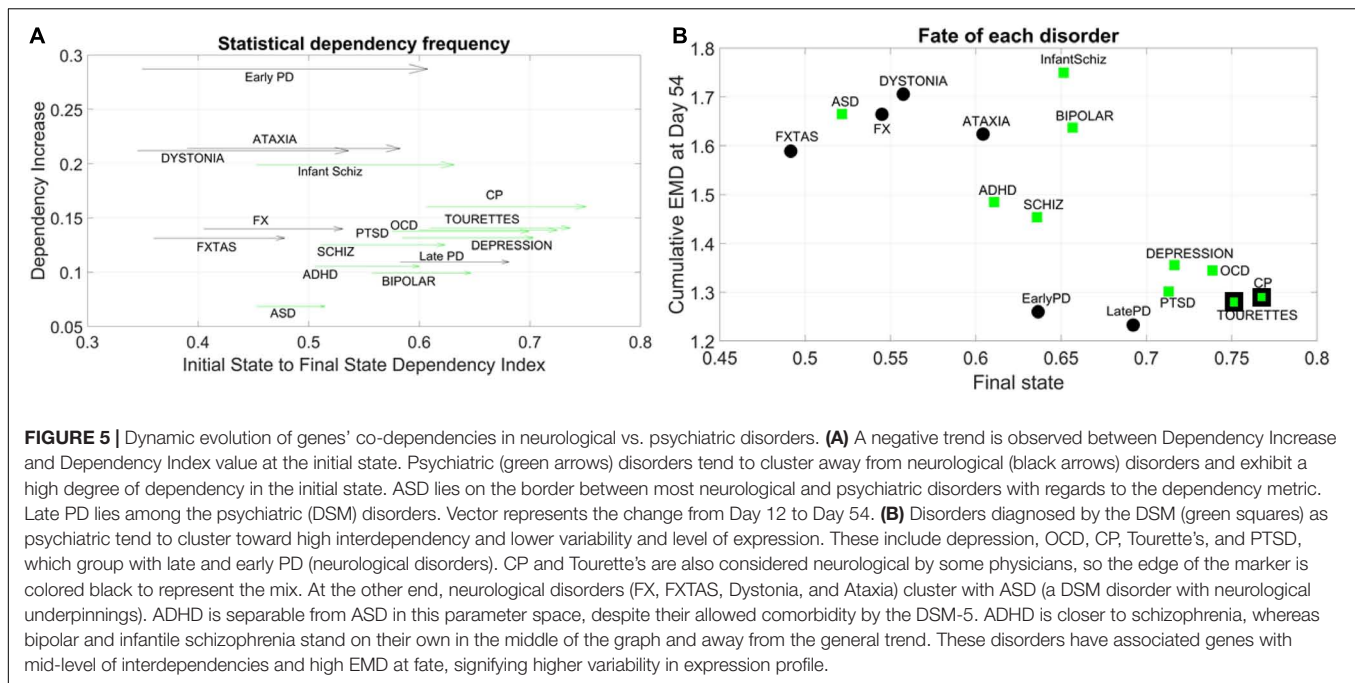
Infantile schizophrenia and bipolar disorder, both DSM disorders, were the exception to this negative trend, as they were both high in EMD expression and dependency index. Further details are shown in **Supplementary Figure 1** comparing the two classes of genes at final fate and state.

## Visualization in Fate Space of Genes Associated With Psychiatric and Neurological Disorders

Using the visualization in **Figure 4C**, we tracked the evolution of the genes associated with the neurological and psychiatric disorders in fate space. We found that they moved from a spread-out configuration in earlier days toward more localized regions along a path of high variability in EMD and an opposite region of low variability in EMD. As explained previously, the EMD measured the change from one frequency histogram (marginal distribution) on gene’s expressions across the cells on one day, to the frequency histogram on the next day. As such, cumulatively they reflect the overall variability in gene expression toward the cells’ fate. We call those with the higher cumulative EMD High Expression Variability (*HEV genes*) and those with the lower cumulative EMD, Low Expression Variability (*LEV genes*).

By day 54, two distinct lobes are obvious in **Figure 6**, which we projected in **Figure 7** along the plane spanned by the first two empirically estimated Gamma moments, mean vs. variance. There we saw the high mixture between the neurological (green) and psychiatric (black) disorders. We then asked if there were distinctions across the genes that we could visualize using the Chow Liu maximal spanning trees, treating their network interactions according to probabilistic graphs.

The results from the visualization prompted us to further examine these intermixed genes and their evolution along the Gamma mean ( $\mu$ ) and Gamma variance ( $\sigma$ ) dimensions of fate space. Projecting these values on a mean, variance parameter plane, clearly showed their evolution and convergence to two distinct lobes in **Figure 7** (Day 54), whereby the lobe of HEV genes with higher variability in EMD quantifying probability transitions, separated from the lobe of LEV genes with lower variability. The composition of these lobes was highly intermixed, with 2,613 genes total, 946 in the lower, line like shaped lobe, and 927 in the upper, curved shaped lobe. That count included all disorders, whereby a gene could be counted multiple times if it was associated with multiple disorders. We then extracted 512 *unique* HEV genes and 927 *unique* LEV genes. These genes might be associated with more than one psychiatric and/or neurological



disorder, or just with one or the other (see **Supplementary Table 1** and expanded **Supplementary Table 2**). **Supplementary Figure 1** shows the dependency indexes and fate across disorders split according to genes in these different lobes. We further analyzed the composition of the two lobes, but first, we had a look at the network analyses.

## Differentiation of Network Evolution of High Expression Variability vs. Low Expression Variability Genes in Psychiatric (DSM) vs. Neurological Disorders

The Chow Liu maximal spanning trees for each of the lobes identified in **Figures 6, 7** were obtained and the degree associated with each gene was quantified to represent in **Figures 8, 9** the network evolution of the HEV and LEV genes, respectively. These graphs show the genes blindly, without the disorders' labels, to give a sense of the differentiation between the two lobes of intermixed genes in both neurological and psychiatric disorders.

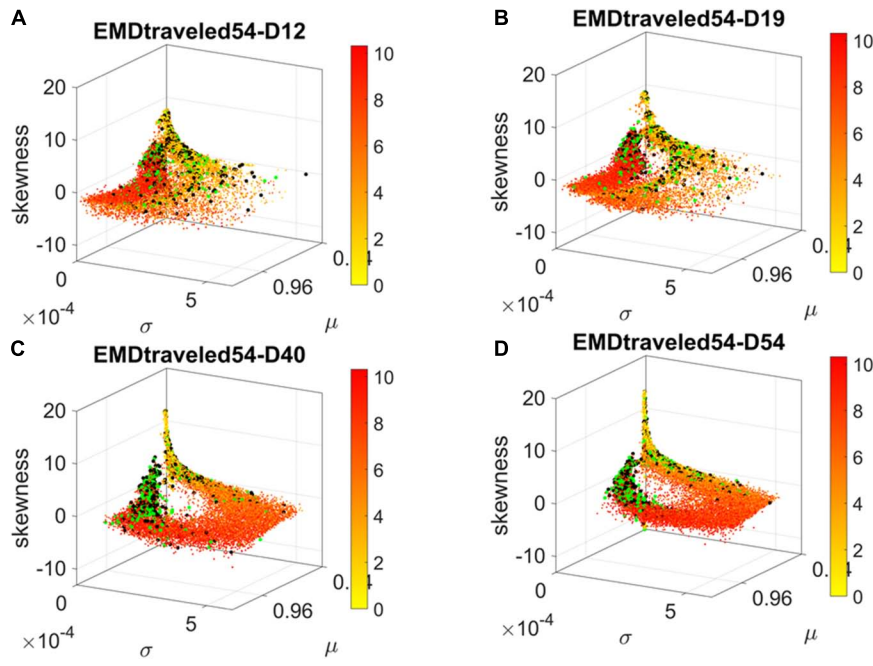
Several important conclusion emerge from these representations of the genes' fate evolution. First, these two lobes have fundamentally different evolution in genes' interactions. Second, the network has a handful of genes with high number of genes connected to it (the network's hubs). These hub genes are such that if a link is disrupted between two of these hubs, the network is disconnected with potential cascade effects (**Figure 8**). In this sense, this is a fragile architecture that remains so at each registered day. Third, days 12 and 54 have fewer hubs than intermediate days. The latter is true in both lobes but more so in the lobe of HEV genes. The lobe of LEV genes has a far more distributed network in intermediate days 19 and 40, suggesting a more robust architecture. This result reveals the importance of

such genes in the overall evolution of the transcriptome toward neuronal states.

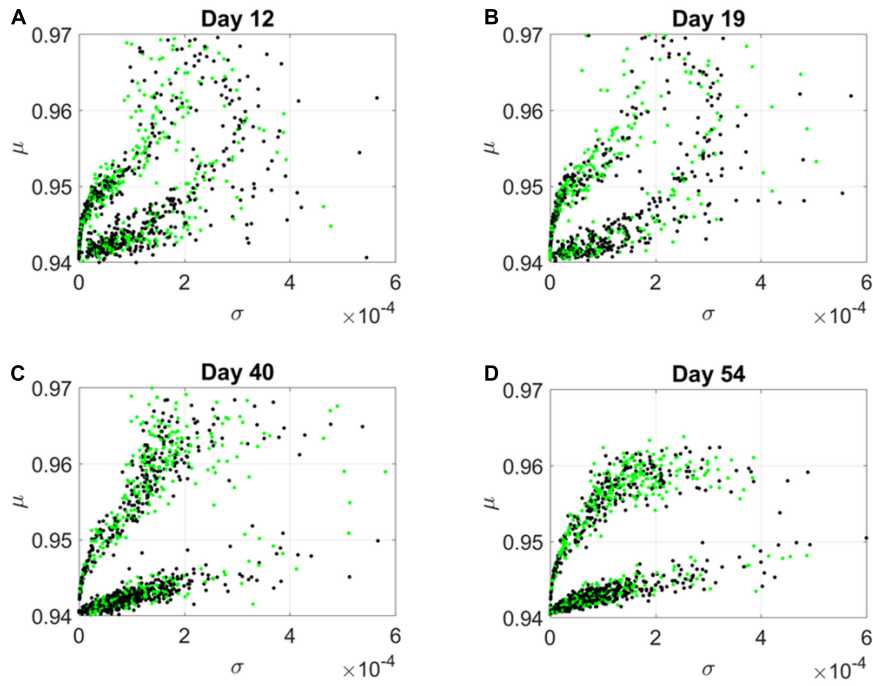
The network patterns prompted investigation of the hubs' evolution across neurological and psychiatric disorders. For each lobe, we built matrices of hub genes along the rows, sorted by graph degree, and disorders across the columns, sorted by neurological (early PD, late PD, Dystonia, Ataxia, FXTAS, and FX) and psychiatric (DSM diagnosed including infantile Schizophrenia, Schizophrenia, ADHD, ASD, Bipolar, PTSD, CP, Depression, OCD, and Tourette's). Each entry was color coded with the node's (gene's) degree (in log scale) to help visualize the colors better, since they ranged non-linearly from 1 to 470. In the case of degree 1, two genes co-depended on each other's expression patterns. In the case of degree 2 or more, the node had an edge with co-dependency with two or more genes. In **Figure 10** we focused on the HEV genes from the curved lobe (with genes colored in green representing those associated with psychiatric disorders and those colored black associated with neurological disorders).

Here the numerator sums over the number of genes with degree above 3 in each of the neurological disorders under consideration and the denominator sums over the number of genes with degree above 3 in each of the psychiatric disorders under consideration. If in a psychiatric disorder (in the denominator) 0 genes participate, the contribution of the denominator is  $e^0 = 1$ . Otherwise, the ratio will reflect the balance of genes in one vs. the other, obtained for HEV and LEV genes separately.

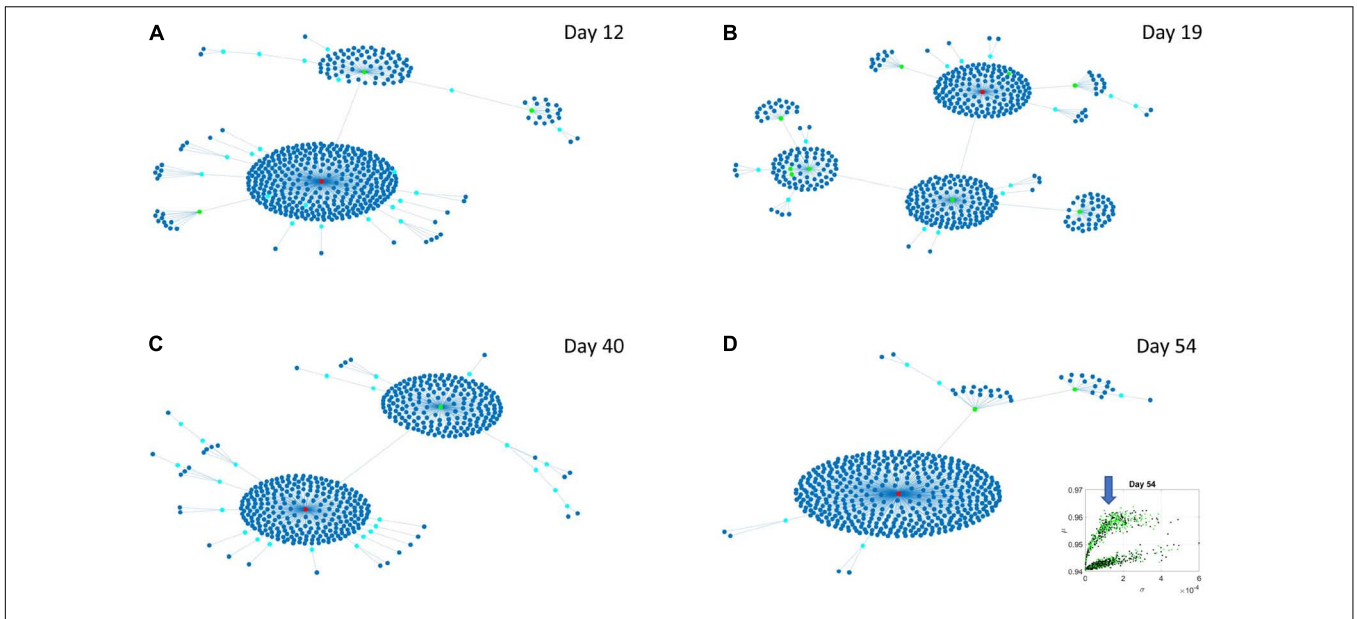
**Figure 10** (HEV genes) and **Figures 11, 12** (LEV genes) depict the matrices whose entry is the scalar value of the ratio (plotted as a color map in logarithmic scale). **Supplementary Figure 2** depicts the HEV genes values and **Supplementary Figures 3, 4** do so for the LEV genes. **Figure 10B** shows the HEV genes lobe



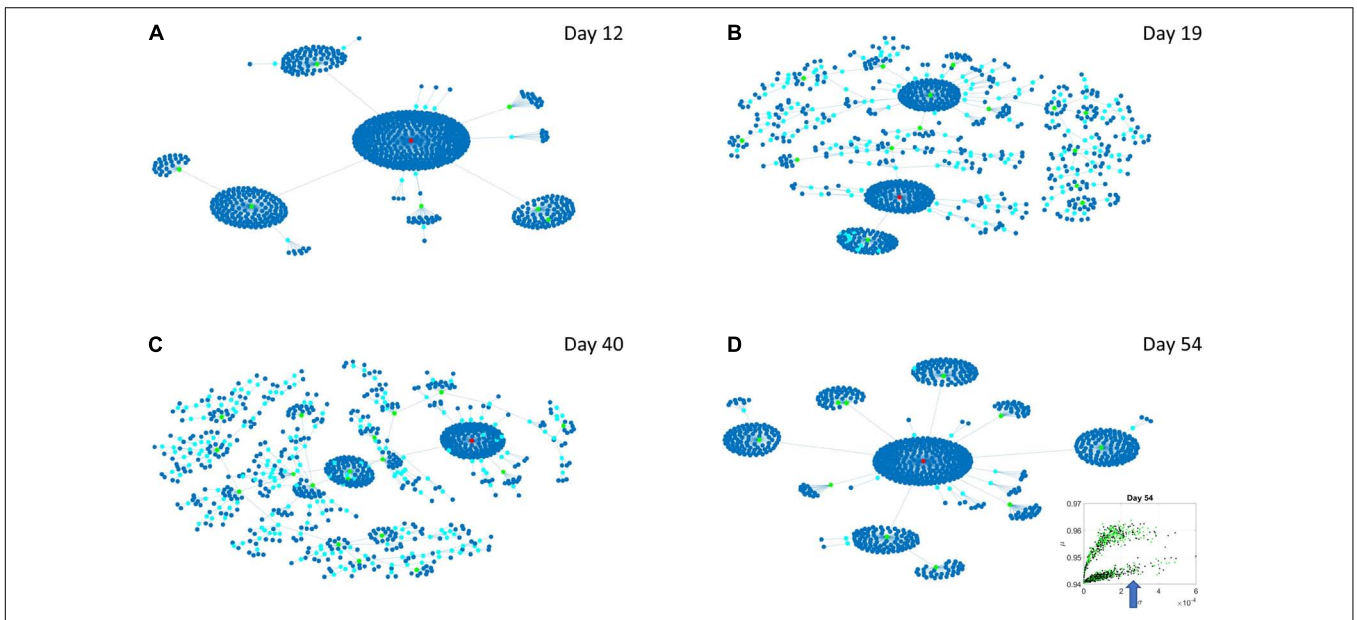
**FIGURE 6 |** Visualizing the trajectories of genes associated with psychiatric (black) and neurological (green) disorders as they evolve in fate space. Color gradient represents the cumulative EMD at day 54, with yellow representing low values owing to low variability and lower expression whereas red represents high variability and higher expression across cells. At day 54, two lobes emerge along the high values space (higher  $\mu$ ) and central tendency (toward 0 skewness) with higher concentration in lower variance regions (along the  $\sigma$  dimension). Notice the high change from Day 19 to Day 40, with clearly two lobes with intermixed genes from both classes of disorders in day 54. **(A)** Day 12. **(B)** Day 19. **(C)** Day 40. **(D)** Day 54.



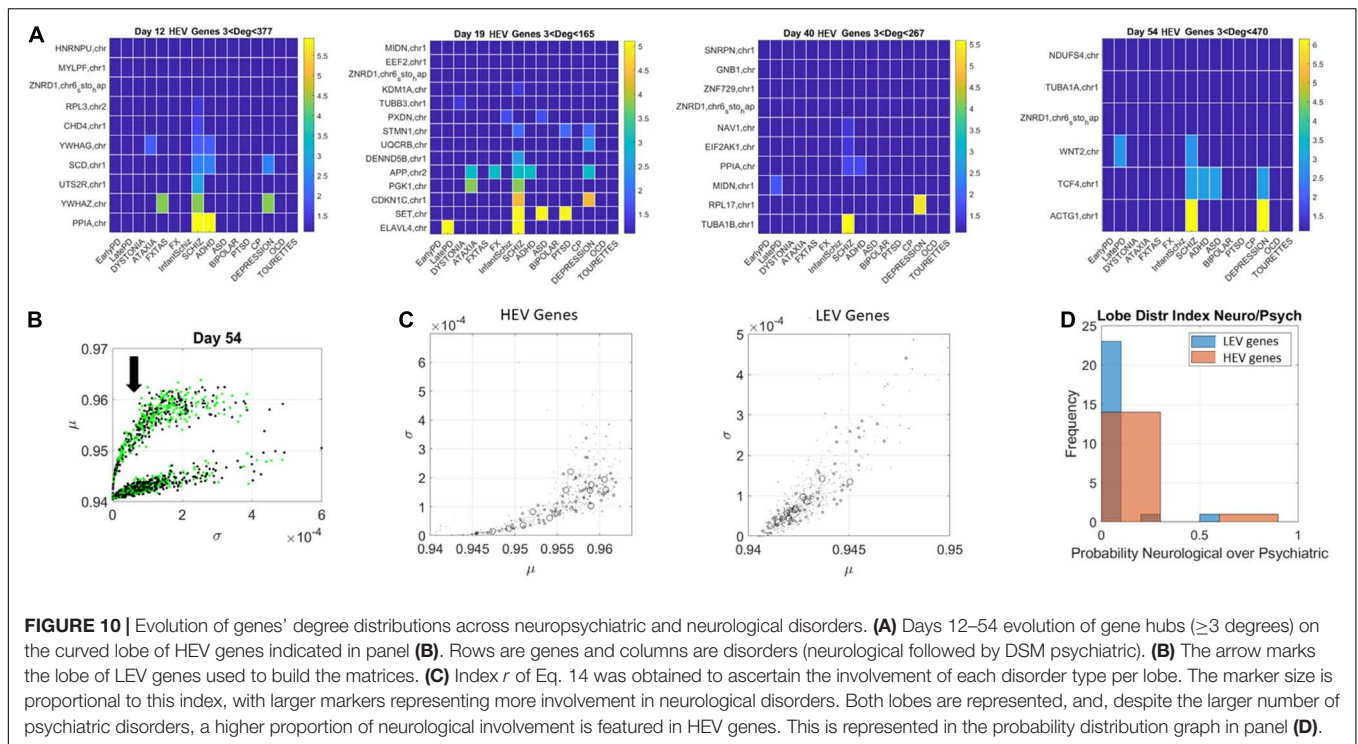
**FIGURE 7 |** Projection of transcriptome genes associated with neurological (green) and psychiatric (black) disorders on the Gamma moments plane spanned by the empirically estimated Gamma  $\mu$  and Gamma  $\sigma$  reveals two classes of genes. At each measurement day the genes move to eventually form two distinct lobes corresponding to the HEV and LEV genes on the curved and line-like scatters, respectively. **(A)** Day 12. **(B)** Day 19. **(C)** Day 40. **(D)** Day 54.



**FIGURE 8 |** Evolution of statistical dependencies of the HEV genes. Hubs with two or more nodes but fewer than 10 are cyan; those with more than 10 degrees but not the maximum are colored green. The maximum degree is colored red. In the inset in the lower right panel, the arrow points at the lobe of HEV genes comprising the dynamically evolving network. The Chow Liu Maximal Spanning Tree represented as an interconnected network evolving from Day 12 to Day 54, for the HEV genes (located on the line like lobe as marked by the lower right-hand inset). The inset in the lower right panel indicates the lobe of HEV genes comprising these dynamically evolving networks. **(A)** The network has two main hubs on Day 12. **(B)** More major hubs appear on Day 19. **(C)** As the network evolves most genes are concentrated once again around two main hubs. **(D)** On Day 54 we have one major hub.



**FIGURE 9 |** Evolution of statistical dependencies of the LEV genes. Hubs with two or more nodes but fewer than 10 are cyan; those with more than 10 degrees but not the maximum are colored green. The maximum degree is colored red. In the inset in the lower right panel, the arrow points at the lobe of LEV genes comprising the dynamically evolving network. The Chow Liu Maximal Spanning Tree represented as an interconnected network evolving from Day 12 to Day 54, for the LEV genes (located on the line like lobe as marked by the lower right-hand inset). The inset in the lower right panel indicates the lobe of LEV genes comprising these dynamically evolving networks. **(A)** On Day 12 the network is centralized with genes tightly concentrated around few major hubs. **(B)** On Day 19 the network is more distributed with a high number of small hubs. **(C)** This behavior persists on Day 40. **(D)** On Day 54 the network is once again centralized but there are more major hubs than on Day 12.



as depicted on **Figures 6, 7** for day 54, whereas **Figure 10C** provides the same plot (rotated) for genes in DisGeNet found on the transcriptome at day 54. In this figure, the size of the marker is proportional to the scalar ratio, reflecting the balance between neurological and psychiatric genes for each class of genes. **Figure 10D** provides the probability graphs depicting the higher probability for HEV genes in neurological conditions. This higher proportion comes despite the far larger number of DisGeNet genes associated with psychiatric disorders like Schizophrenia [as captured by the colormap matrices in 10A (HEV genes) and 11–12 (LEV genes)].

Similarly, for the LEV genes, we plotted the color map matrices involving disorders along the columns and genes along the rows. To aid with visibility, these are split between **Figure 11** for Days 12 and 19, and **Figure 12** for Days 40 and 54. In days 19 and 40, more genes partake as hubs and sub-hubs, with a more distributed network topology than in the case of HEV genes (as shown by the networks in **Figures 8, 9** for HEV and LEV genes, respectively).

We saw that schizophrenia (being the disorder associated with the highest number of genes across these disorders) was the one with the highest number of hubs. We also observed the involvement of a hub across multiple disorders of the neurological class and of the psychiatric class. Furthermore, we noted that psychiatric disorders in general spanned more hubs, and these hubs had a higher degree than neurological disorders. Hubs with degree of 3 and above that were common to both classes of disorders and / or in only one class abounded and are listed in **Supplementary Table 1** and in the expanded **Supplementary Table 2** listing also function and other information about the genes.

To further investigate the balance between HEV and LEV genes in neurological vs. psychiatric disorders, we quantified the disease ratio,

$$r = \frac{\sum_{genes} NeuroDegree}{\sum_{genes} PsychDegree} \quad (14)$$

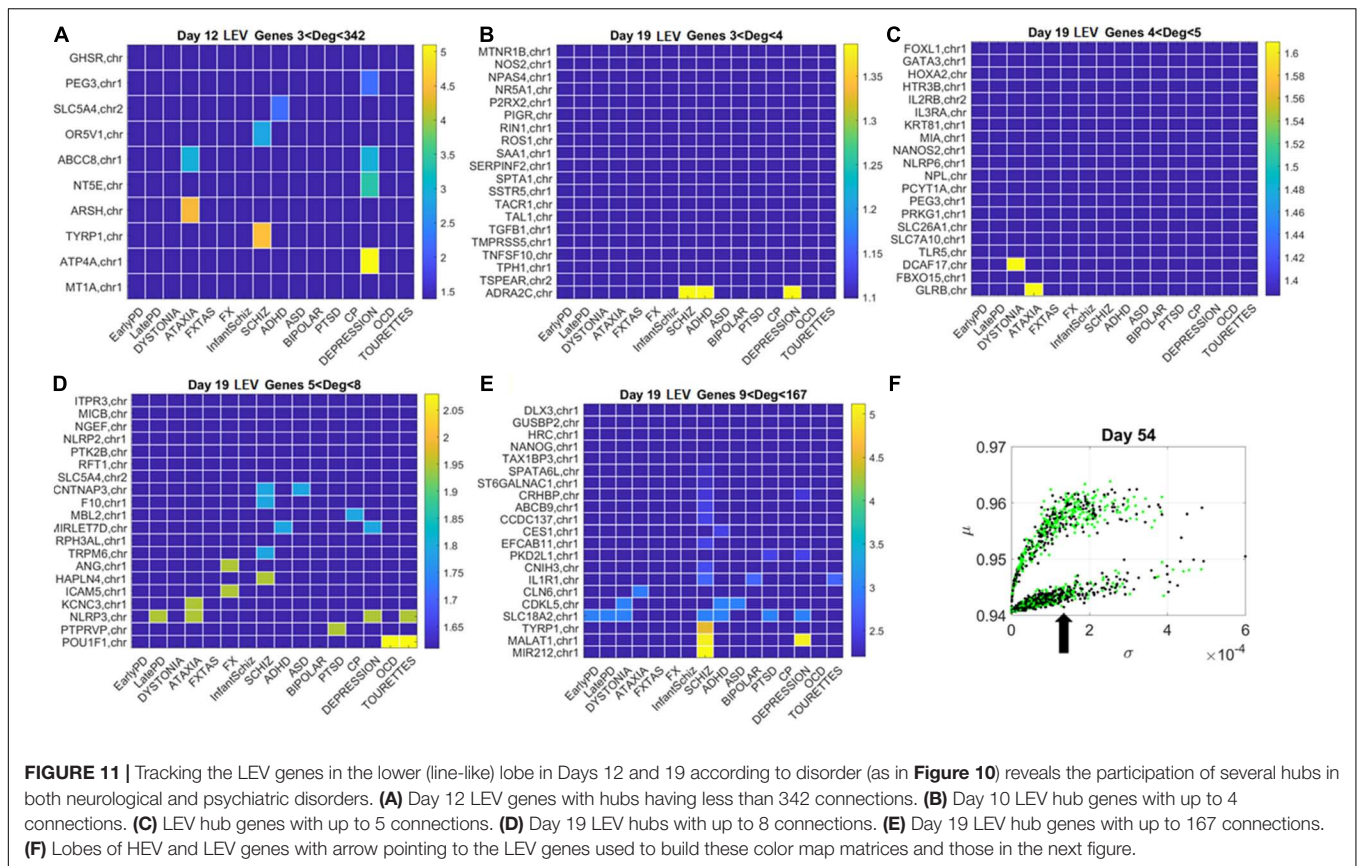
### Results From Binary Barcode (State Space) and Cells' Subtypes

**Figure 13** shows representative results of tracking the gene's state according to a binary code that sets the state ON if the node's degree at a given day is above the average degree of the network that day, and OFF otherwise. With four readings we have  $2^4$ , 16 possible states, and each gene is assigned one possible state according to its network dynamics. In this figure (for simplicity) we plot two examples of genes with state [1111] remaining each day above the average degree of the network, taken across all nodes. Likewise, the example of [0100] is of a gene that only raises above the average degree of the network in day 19. The other configurations (dashed trajectories) on each plot reflect the trajectories of other genes in other binary state vectors.

Using this information, we could build clusters of the cells that primarily had genes in a binary state. The use of both the fate and state dynamics of the gene enabled the tracking of the stochastic activity over time.

### Further Similarities and Differences Between Neurological and Psychiatric Disorders

Besides tracking genes expression and their variability in fate space along with the binary ON/OFF states of each gene and



their involvement in neurological vs. psychiatric conditions, we selected the following clusters of tissues to evaluate genes' involvement:

- Cortical Regions: Cortex, Frontal Cortex.
- Neuromotor System: Caudate Basal Ganglia, Cerebellar Hemisphere, Cerebellum, Nucleus Accumbens, Putamen Basal Ganglia, Substantia Nigra.
- Limbic System: Amygdala, Anterior Cingulate Cortex, Hippocampus, Hypothalamus.
- Spinal Column: Spinal Cord Cervical C1, Nerve Tibial.
- Glands: Adrenal and Pituitary gland.

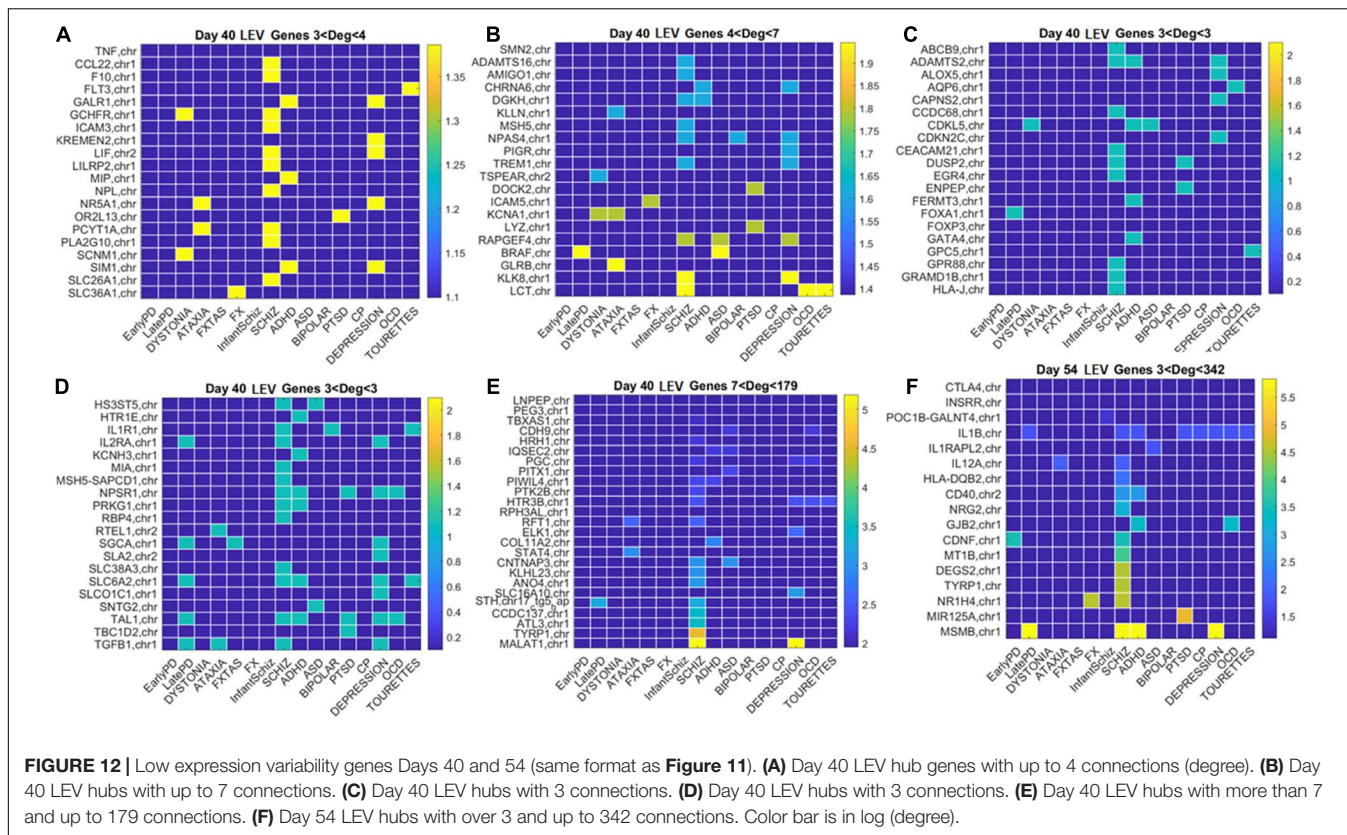
We found statistically significant differences at the alpha 0.05 level ( $p < 0.04$ ,  $t$ -test). **Figure 14** shows the bar plots comparing the outcomes for HEV and LEV genes for psychiatric (red) and neurological (blue) disorders.

The psychiatric disorders included ADHD, ASD, Schizophrenia, Bipolar, Depression, Infantile Schizophrenia, PTSD, and OCD. The neurological disorders included FXTAS, Early PD, Late PD, Ataxia, FX, Dystonia and CP, and Tourette's, given the involvement of the neuromotor systems in the last two disorders. Maximal differences in percentile change and in the reshuffling of disorders (which we plotted sorted by percentile in **Figure 14**) accounted as well for statistically significant differences between the tissues for the HEV vs. LEV genes ( $p < 0.04$ ,  $t$ -test).

## DISCUSSION

This paper uses genomic information to reframe a recently revived debate on the possible differentiation between neurological and psychiatric disorders. We reframed the question by addressing whether, despite a shared genomic pool, psychiatric and neurological disorders could be differentiated at very early embryonic stages. To pursue our question, we took advantage of the validation strategies used by Yao et al. (2017) (i.e., the fact that they compared their two-dimensional *in vitro* model to primary tissues from atlas data and cortical cells from mid-gestation human fetal embryos). Their work generated a transcriptome-based lineage that allows for studies of human brain development and for the modeling of human neurodevelopmental disorders.

With these results in mind, we here developed and implemented a three-level approach that interrogated the early development human transcriptome trajectories of hundreds of hESCs as they reached neuronal state. Each level of inquiry offered new insights on the complex genetic origins of psychiatric and neurological disorders, highlighting fundamental differences between the two types of disorders. We uncovered and characterized two classes of genes with essentially different dynamics (distinct ON/OFF states) and fate (cumulative expression variability) featured throughout differentiation. Furthermore, using these classes of genes, we pointed at commonalities between the motor and limbic



systems for one class (but not the other) that could possibly explain the current confounds in observational criteria. These analyses provide evidence supporting the notion that psychiatric disorders have substantial neurological underpinnings and yet their associated genes’ network interdependencies are significantly different at the early embryonic stages of cell differentiation. We next discuss each level separately.

### Level 1: Marginal Distributions

Traditionally, transcriptome interrogation aims at uncovering different classes of cells with some genomic composition. This general approach tends to do away with genes that have low variability or asynchronous behavior, i.e., they may be turned OFF in the initial stages, or have such low expression that their contribution is presumed negligible. We thought differently here and, instead of first trying to uncover cells’ classes, we transposed the cells × genes matrix and expressed each gene as a function of the cells’ expressions. Then the question was, for each gene, how was the gene’s expression across cells cumulatively contributing to the final neuronal fate. Furthermore, how was the gene’s state evolving across these different readings? We reasoned that some registered cells might not be the same from day to day, yet the expression of the genes would change across the transcriptome in some way that would lead to self-emergent patterns, found without discarding any genes. For each gene we then obtained the marginal probability distribution of its expression to neuronal fate and measured across days the departure in expression variations, using the EMD metric

appropriate to quantify differences or similarities between such frequency histograms.

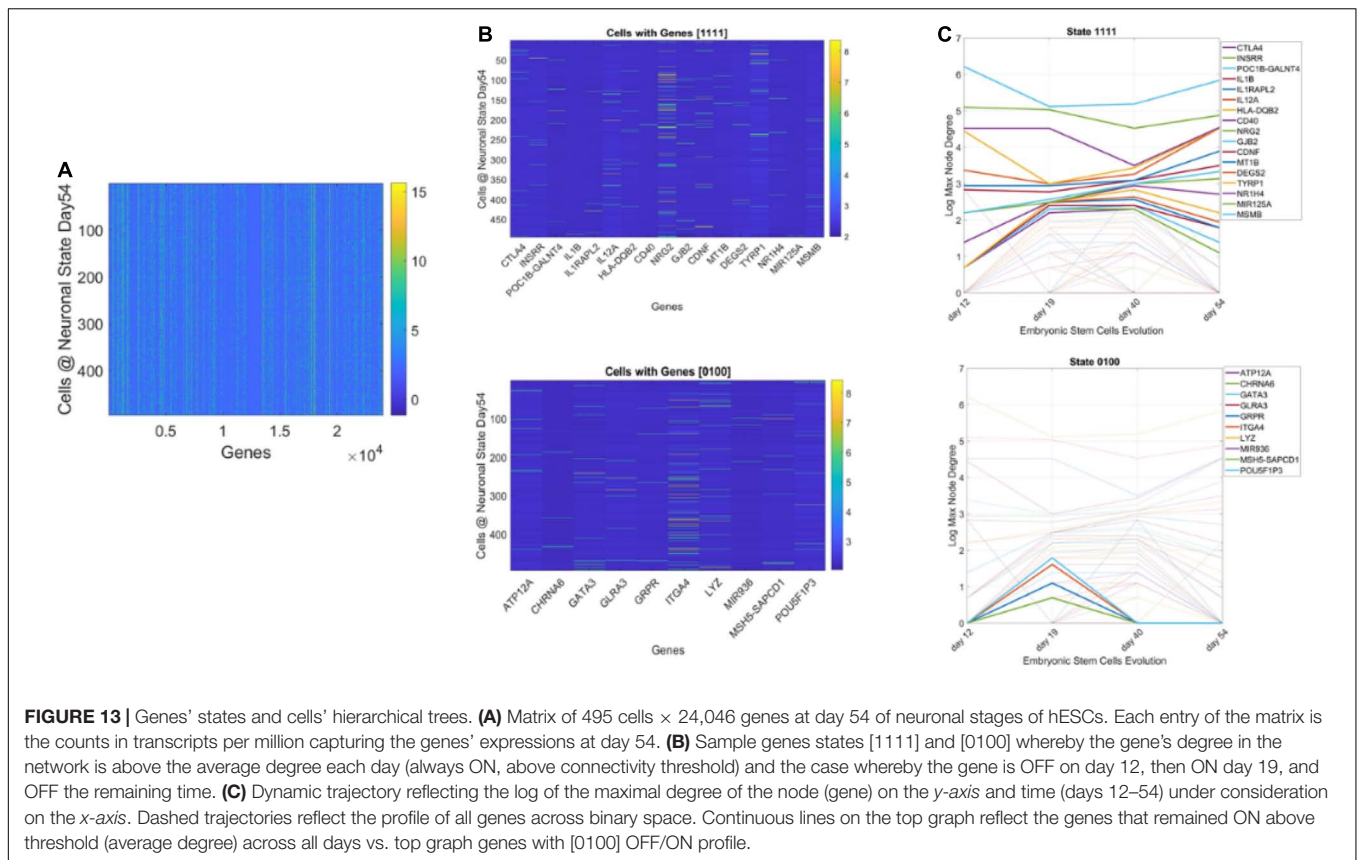
Tracking this information without discarding any genes allowed us to visualize the genes associated with psychiatric and neurological disorders embedded in the full transcriptome. The new visualization tool revealed *two fundamental subtypes of genes* (as depicted in the two lobes of **Figure 7**). Both lobes had a mixed composition of genes associated with psychiatric (according to the DSM-5) and neurological disorders. The question that emerged then was, what contribution was each set of genes (neurological vs. psychiatric) making to each lobe? We will defer that question to the third level of inquiry.

One lobe is characterized by genes of highly varying expression and high cumulative EMD whereas the other contains genes of low expression variability and low cumulative EMD. Here it may be worthwhile mentioning that traditional methods such as *PCA* and *t-SNE* would have likely missed the LEV genes of the second lobe, with lower expression and lower variability. And as we will see at the third level of inquiry, that would have missed an opportunity to capture the real evolution of these genes from the vantage point of probabilistic nodes interacting across a network.

### Level 2: Co-dependent Genes

Moving on to a higher level of complexity, we studied the pairwise interactions between genes by focusing on the joint probability distributions of all possible pairs for each distinct group of genes associated with the various psychiatric and neurological





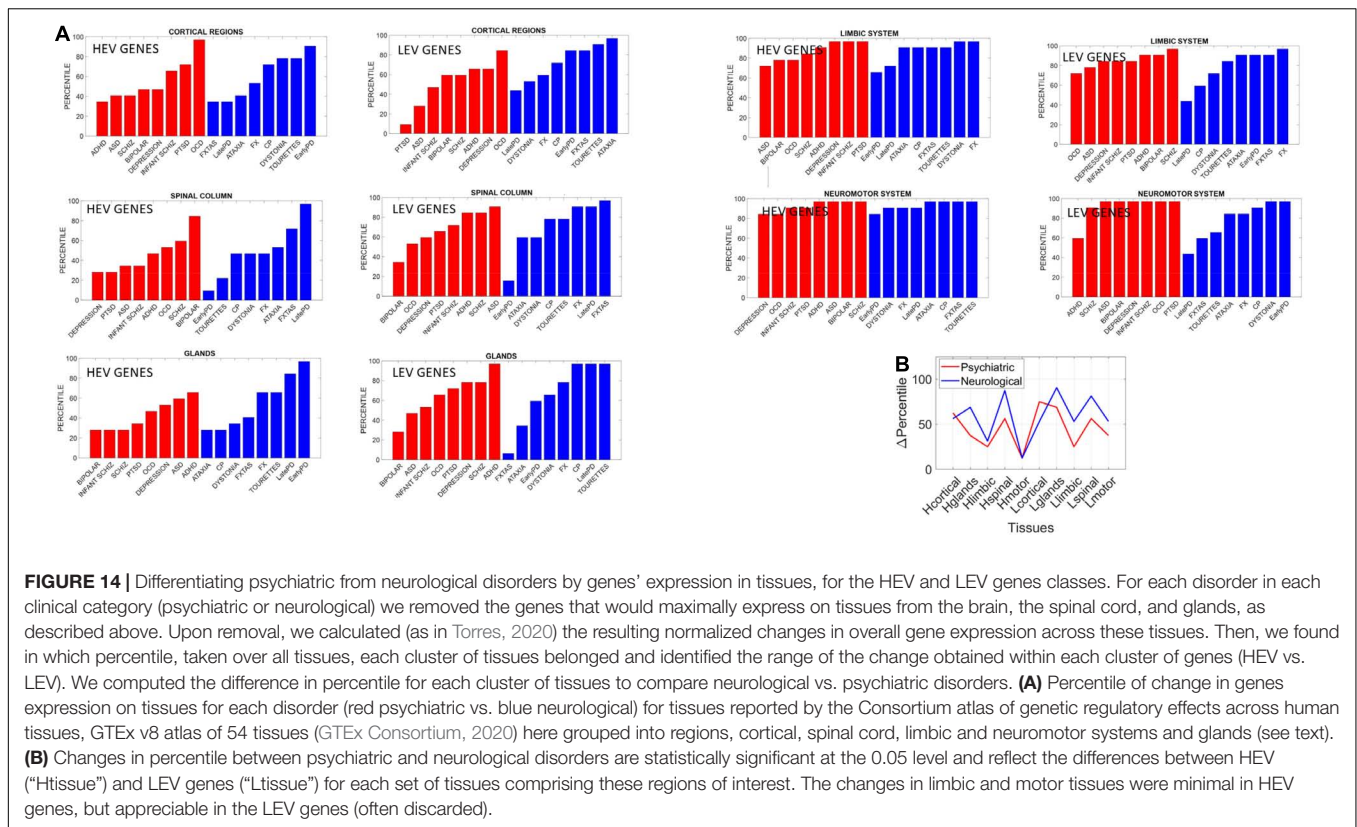
disorders, according to sources found in the DisGeNet portal. By employing a robust statistical independency test, we were able to quantify the average degree of pairwise dependencies in each network of gene expressions. The initial degree of interdependence between genes associated with different pathologies (Figure 5A) allowed us to differentiate between neurological and psychiatric disorders.

Two main clusters appeared in this vector field, representing the increase in dependency from the initial to the final states. One cluster boasting a higher dependency increase was primarily neurological: early onset PD, Ataxia, Dystonia, FX and FXTAS. Yet, several DSM (psychiatric) disorders appeared at that level as well. These included disorders that are detectable in infancy, such as infantile schizophrenia, CP, OCD, and Tourette's. They share a highly compromised somatic-sensory-motor system and profound issues with the limbic system that without a doubt will impact the overall neurodevelopment of the individual. This pool of genes with high dependency gains across early embryonic stages of neuronal cell differentiation suggests rather early origins of such disorders and the highly interconnected evolution of their associated genes. They also showed a larger rate of change from the initial to the final state.

At a lower level of dependency increase, we saw mostly DSM psychiatric disorders, except for late onset PD. This is interesting considering the high incidence of dementia, hallucinations, and other altered mental states in late PD and other related tauopathies (Koga et al., 2022; Tu et al., 2022). These rates of

change were more modest than those in the other group of disorders with a higher dependency increase. In particular, ASD, which lies approximately midway along this vector field, had the lowest dependency increase, signaling an altogether different signature of genes' co-dependency during early embryonic stages of neuronal differentiation.

Indeed, ASD seems to lie on the border between the class of neurological and psychiatric disorders, which confirms at the genetic level that the spectrum of autism comprises pathologies of the nervous system that underlie its phenotypic conceptualization as a behavioral/mental disorder. This is supported by extended research showing biorhythmic patterns in autism with a unique signature of noise-to-signal ratio derived from fluctuations in signal amplitudes and timing. This motor code is bound to impact the kinesthetic reafferent feedback from the periphery to centers of central control (Torres et al., 2013a, 2020; Torres and Denisova, 2016; Wu et al., 2018; Torres, 2021). Here we observed the origins of such departures at these early embryonic stages of neuronal cells' differentiation, whereby the genes associated to ASD manifested the smallest shift in dependency index value from the initial to the final state, accompanied by the smallest increase in genes' statistical dependency. The change in dependency index (i.e., the range of values where the arrow denoting rate of change lands) overlapped with those ranges in FX, FXTAS, Ataxia, Dystonia, and Early PD along the neurological disorders, and with infant Schizophrenia, along the psychiatric ones. This is interesting in view of recent



work examining digital biomarkers of FX carriers who, despite their young age, manifested gait patterns present in Ataxia and FXTAS much like participants in the autism spectrum did (Torres et al., 2020; Bermperidis et al., 2021). In this recent work, according to causal stochastic analyses, the kinesthetic feedback loops estimated from their gait largely departed (in both ASD and young FX carriers) from the neurotypical age-matched controls. They resembled instead the gait of patients with PD (Bermperidis et al., 2021). Our results here confirmed that the presence or absence of a disorder was not due to the mutation of a specific gene but rather resulted from the degree of co-expression among multiple genes. Therefore, in the context of the evolving human transcriptome, what truly separates neurological from psychiatric disorders is that the latter show much higher complexity of co-expression and interconnectedness in the early stages of cell differentiation, as compared to the genes associated with the former.

This hypothesis is reinforced if we observe the relative positions on this plane of Infant Schizophrenia and Schizophrenia, Schizophrenia and Bipolar Disorder, Tourette's and OCD, PTSD and Depressions as well as ASD and ADHD, and ASD and Infant Schizophrenia. Once again, we have a transition in the initial degree of dependency from what characterizes Infant Schizophrenia as a neurodevelopmental disorder to what characterizes Schizophrenia as a disorder appearing in early adulthood. Also, OCD and Tourette's have a high comorbidity, and symptoms of the latter may appear in the symptomatology of OCD. The same is true for Depression and

PTSD, whereas both Schizophrenia and Bipolar Disorder belong in the psychotic spectrum of psychiatric diagnoses, according to the DSM-5. Hence, the proven clinical proximity in all three pairs recapitulates here in their proximity in our parameter space. Interestingly, ASD was once labeled as "Infantile Schizophrenia," a clinical labeling that receives support in our parameter plane: ASD and Infantile Schizophrenia have nearly the same initial degree of dependency in our parameter space. Finally, Tourette's is clustered among the psychiatric disorders. Indeed, there is ample debate on whether Tourette's is a psychiatric or a neurological disorder, despite generally being classified as the latter (Sandor, 1993; Kerbeshian et al., 1995; Brovedani and Masi, 2000).

When examining the cumulative changes in expression variability according to the EMD, a negative trend emerges in both types of disorders. The higher the genes' variability accumulated toward Day 54, the lower the dependency index in this final state. However, ASD appears among the neurological disorders and Early and late PD appears among the psychiatric (DSM) ones (near to PTSD, Depression, OCD, CP, and Tourette's). Furthermore, ADHD and Schizophrenia lie midway of this scatter and Infantile Schizophrenia and Bipolar disorders depart from the psychiatric group along the axis of variability, i.e., their average cumulative EMD expressions are among the highest levels, along with those of the neurological disorders, at Day 54.

From these patterns, it may be possible to perceive that networks that reach a highly interconnected and complex state are characterized by genes that have more constancy in their

statistical behavior. We revisit this proposition shortly, at the third level of inquiry. We reasoned here that, intuitively, this would make sense, since highly interconnected networks could dictate a gene's behaviors in a distributed way, whereas more disjointed networks would allow genes to behave in a more independent manner.

Our second level of inquiry considers pairs of genes, isolated from the rest of the network, and then integrates over all pairs to derive the degree of interdependency. We next explore a full probabilistic graphical model and visualize the global behavior of these genes. We characterize the topology of this network and identify the importance and role of different genes in the evolution of the network and subsequently on the origins of different disorders. To that end, we consider the multivariate probabilistic behavior of the two different subtypes of genes that we discovered at the first level of analysis.

### Level 3: Genes' Networks

By employing factorization techniques and graphical modeling, we were able to capture the evolutions of the networks of HEV genes (**Figure 8**) and LEV genes (**Figure 9**). In both graphical trees, we observed a hierarchical structure, with genes that were central (hubs) and associated with many other genes as well as genes that were leaf nodes. One key difference between HEV and LEV genes was that the latter have a less hierarchical organization on Days 19 and 40, with many small hubs and "clouds" of genes forming around the dominant hubs of the network. Note that the topology of the networks implies fragility, since removing a central hub or removing edges that connect large hubs would result in disconnected graphs, with the network of HEV genes being the most fragile of the two.

We then cataloged the degrees of genes that were significant in each of the eight networks (two networks  $\times$  4 days for each genes' class) and belonged to different neurological and psychiatric disorders (see **Figures 10–12**). Despite a vast majority of high degree genes being consistently associated with psychiatric disorders such as Schizophrenia (owing to the very high number of associated genes reported in DisGeNet), we found a far higher proportion of neurological involvement in HEV genes (**Figures 10A,D** showing the probability distributions along with **Supplementary Figures 2–4**). This result, along with those in **Figures 11, 12** for the LEV genes, pointed to a degree of overlapping of genes associated with neurological disorders with those associated with psychiatric disorders. This is captured as well on **Supplementary Table 1**, which we expanded **Supplementary Table 2**, to catalog the genes' function, location, and phenotypes. The results also unambiguously separated genes involved in neurological disorders from those associated with psychiatric disorders in that the former were probabilistically more HEV (**Figure 10D**). In contrast, the latter tended to be (probabilistically) LEV, yet forming more robust and distributed networks. From the network analyses and the index ratio quantifying neurological over psychiatric predominance, we concluded that, probabilistically, underlying psychiatric disorders have more LEV genes and underlying neurological disorders have more HEV genes.

### The Level of Average Genes' Expression on Tissues

The fundamental differences quantified between psychiatric and neurological disorders using the different lobes of HEV vs. LEV genes extended to the tissues, as we probed these genes' expressions on the 54 tissues from GTEx. Here we grouped tissues into brain regions (cortical, motor-subcortical and limbic), the spinal cord, and glands (see main text for details) and measured the percentile change in HEV vs. LEV genes upon removal of those genes in each of the disorders under consideration. We then grouped these disorders according to psychiatric and neurological clinical classification and compared the change in percentile between the two types of disorders.

Our analysis showed that, across different groups of tissues that typically serve different roles in the autonomic, central, and peripheral nervous system, fundamental differences emerged between the two types of disorders under consideration. The genes associated with psychiatric disorders expressed on these tissues differently than they did in neurological disorders. These differences were statistically significant between HEV and LEV genes in general. They were also statistically significant when considering the genes as part of psychiatric or neurological disorders. To that end we pooled the changes in percentile across all tissues and genes' type of each disorder class (psychiatric or neurological) and found that the HEV vs. LEV genes classification served to separate psychiatric from neurological disorders. Furthermore, motor, and limbic regions were minimally different in HEV genes (relative to LEV genes) when comparing psychiatric to neurological disorders (**Figure 14B**). This suggests that LEV genes, traditionally discarded, contribute to such statistically significant differences between psychiatric and neurological disorders. This result, paired with the probabilistically higher prevalence of neurologically associated genes across the HEV lobe and their presence in psychiatric disorders, supports the notion of motor involvement in mental pathologies. It is in this sense that one could argue that psychiatric disorders are also neurological disorders.

A main corollary of these results is that the LEV genes, which under traditional methods are likely discarded and excluded from the analyses, make an important contribution to the distinction between psychiatric and neurological disorders. This can be appreciated in **Figure 14B** whereby HEV genes (likely the ones entering the analyses under traditional techniques) do not separate motor and limbic systems between the two disorder types. It is the LEV genes that do so in the motor and limbic systems, and in other tissues as well. Further development of new analytical techniques that also include these genes with lower expression variability and asynchronous ON/OFF states may open new lines of inquiry across diseases in general.

### Considering the Gene's State Through a Binary Barcode

Interrogating the fate of the genes gave us a sense of ways to automatically cluster groups of genes according to the evolution of their expression variability. But, what about the changes in gene's state? Once we reached the third level of inquiry and

determined the gene's degree of codependency at the network level, it was possible to examine the average degree of the gene to determine its hub activity level as above or below that average threshold (on a logarithmic scale to capture the non-uniform degree distribution). In 4 days of measurements (in this case), we found genes that were always ON [1 1 1 1] in their hub networking role. These stood in contrast to those that were always OFF [0 0 0 0] in this role. Then, we found genes in other categories, thus building a barcode of binary states that allowed further examination of the genes embedded in the full transcriptome. Here we studied these pools of genes associated with neurological and psychiatric disorders, but the same type of method could be employed to interrogate the network's state dynamics of genes associated with other disorders, with different frequency of recordings (here  $2^4$  in 4 days of recordings, yielding 16 types that resulted in different cell classes expressing those main types). However, other refinements of this method might offer new dynamics information with higher granularity of cell's state that includes the low- and odd-varying genes with asynchronous states (such as the class of LEV genes uncovered here). In this sense, further work is warranted to formalize a new embedding algorithm that considers the full transcriptome fate and state code, to reveal topological invariants of the genes' classes associated with known diseases.

## Recapitulating Phenotypical Information Cataloged From Multiple Sources

We took our investigation one step further and compiled information from the OMIM online source for genes of degree  $>3$  of both gene lobes (HEV and LEV). Accordingly, we examined three categories of genes that, according to their association with both neurological and psychiatric disorders (overlapping roles), were associated with neurological disorders only and those associated with psychiatric disorders only. The "Neurological only" category consisted of only two such genes with degree higher than 3. Information about these classes of genes can be found in the **Supplementary Table 1** and in the expanded version of this table, including the inheritance of the identified genes and the functional properties of the proteins they encode in the **Supplementary Material**.

Both sets of genes associated only with psychiatric disorders and those associated with both the neurological and psychiatric ones were found to play critical roles in the developing nervous systems overall, fetal brain development, neurogenesis, neuronal differentiation, neuromodulation, synaptic organizations, healthy function of the senses, cortical development, and neuronal migration in the context of development. The psychiatric only-type genes were mainly associated with serotonin (5-HT), glutamate (NMDA), and nicotinic acetylcholine receptors (nAChRs), whereas the neurological and psychiatric-type genes were associated with adrenergic (norepinephrine) and dopaminergic pathways. In both groups, specific genes were important for the GABAergic system.

In an interesting outcome (summarized in **Supplementary Figure 5**), genes with a high network degree (large hubs) turned out to be crucial to the immune system. These genes are

associated with autoimmune disorders and neurodegeneration, synthesis of growth factors, cancer and metastasis, inflammatory responses, and allergies. Notable cases are the HEV gene ELAVL4, with a network degree of 165. This hub is associated with Late Onset Parkinson's disease and Schizophrenia and is related to paraneoplastic neurological disorders (PND) and autoimmune neuronal degeneration. Another gene, LIF, in the LEV genes lobe, with a network degree of 4, has been hypothesized to play a functional role at the interface between the immune system and the nervous system. Here we recapitulated its association with Schizophrenia and Depression.

According to the OMIM literature, many of the identified high-degree genes (the hubs) play key roles in basic molecular and cellular functions, such as signal transduction, ATP synthesis, mitosis, cell-to-cell adhesion, intracellular signaling pathways, differentiation, proliferation, and transcription regulation. These large hubs also participate in these processes by influencing other genes' functions (as predicted by the network's topology) implicated in the formation of key components of the cell, such as the cytoskeleton and the extracellular matrix. These hub genes are crucial to the survival of the eucaryotic cell.

*These findings imply that the genes associated with neurological disorders are no more fundamental than the genes associated with psychiatric disorders or that are at the intersection of both disorders.* According to the transcriptome data used here, validated using primary tissues from atlas data and cortical cells from mid-gestation human fetal embryos, and to the analyses that we performed, the origins of both types of disorders can be traced back to the embryonic stages of differentiation and development. These involve the emergence of fundamental neural pathways and general biochemical cascades that characterize eucaryotic cell life. Moreover, in both types of disorders, key genes support the role and interaction of the immune system with the developing nervous system. The interaction between the two sub-systems is being explored and investigated by various researchers (Ashwood and Van de Water, 2004; Ashwood et al., 2006; Michel et al., 2012; Meltzer and Van de Water, 2017), and their findings will shed light not only on the mechanisms of emergence of neurodevelopmental and neurological disorders but also on what the DSM-5 characterizes as "mental disorders."

Amidst such heated debate on differentiating between neurological and psychiatric pathologies, perhaps, as synthesized by the OMIM literature behind their associated genes in our networks, a fundamental distinction is delineated by the neurotransmitter receptors and pathways linked to these genes. For example, the degeneration of dopaminergic pathways in the striatum is responsible for Parkinson's disease-associated tremor, which is, at some point of the disorder's progression, observable to the naked eye, thus shifting the focus to the "neurological" nature of PD (while ignoring the progressive dementia associated with it). In contrast, low serotonin in Depression leads to a "psychiatric" phenotype that sidelines motor control issues in these patients. On the other hand, high dopamine levels in Schizophrenia seem characteristic of this disorder, which despite being labeled "psychiatric" has a definite motor component (Rogers, 1992; Walther and Strik, 2012;

Nguyen et al., 2016; Slowinski et al., 2017; Walther and Mittal, 2017; Walther et al., 2020, 2022).

## Possible Utility of Reframing the Question of Psychiatric vs. Neurological Disorders From the Genomics Standpoint

Two important theoretical constructs could be proposed to further disentangle these disorders at the level of modeling system's behaviors. One theoretical construct would have to borrow mechanisms from the immune/autoimmune systems to frame models of possible mechanisms. Given the synthesis of OMIM information, this line of inquiry will be important in future computational work.

Another avenue of theoretical inquiry explaining mechanisms to differentiate neurological from psychiatric disorders at the behavioral level would be related to the general framework comprising internal models of action (IMA) in the field of neuromotor control (Kawato and Wolpert, 1998; Wolpert and Kawato, 1998). We refer specifically to the “*principle of reafference*” (Von Holst and Mittelstaedt, 1950) and related computational modeling. According to this basic principle, every time a movement is initiated by the nervous system, information is sent to the motor system and a copy of the signal is created, known as the *effeference copy*. This enables the CNS to distinguish sensory signals stemming from (exogenous) external environmental factors from (endogenous) sensory signals coming from the body's own actions (Jekely et al., 2021). According to the IMA, the *effeference copy* is provided as input to a forward model, which predicts the sensory consequence of a motor command and measures the error between desired and attained outcome (Kawato and Wolpert, 1998; Wolpert and Kawato, 1998). Although IMA focuses only on error-correction and targeted-directed actions, complex movements are richly layered (Brincker and Torres, 2018). As such, *endo-afference* can be further separated into at least two components based on different classes of movements (Torres, 2011). One type of *endo-afference* is classically associated with voluntary actions, i.e., those deliberately aimed at a goal and operating under an error correction code (Kawato and Wolpert, 1998; Wolpert and Kawato, 1998). Another type of *endo-afference*, however, is associated with spontaneous or incidental actions, encompassing signals that transmit information about contextual variations associated with exploratory learning (Torres, 2013; Vaskevich and Torres, 2022). Reafferent signals also include pain and temperature afference, a far more complex and elusive code that needs to be distinguished from the motor code in current biosignals' analytics, e.g., as revealed in Elsayed (2021) and Ryu et al. (2021).

The principle of reafference allows us to distinguish between different levels of *mental intent* (Torres, 2013; Torres et al., 2013b; Choi and Torres, 2014; Ryu and Torres, 2020) and *physical volition* (Torres and Lande, 2015; Nguyen et al., 2016; Torres, 2016, 2017). This distinction can be conceptually mapped onto psychiatric and neurological issues,

respectively, to inform hypothesis testing and theoretical modeling, possibly expanding criteria beyond subjective observation and opinion. However, we feel that the theoretical mechanisms stemming from immune/autoimmune systems will non-trivially add to our understanding of genomic differentiation between these classes of disorders. As such, they should be incorporated in a new internal model framework amenable to address different genes' classes analogous to deliberate vs. spontaneous or error-corrective vs. exploratory modes of neuromotor control and learning, respectively (Torres, 2011; Vaskevich and Torres, 2022). Here understanding recurrent loops of genes modulating other genes will be critical to forecast and track the onset timing of these disorders.

In summary, we have demonstrated at the level of hESCs that there are fundamental differences between psychiatric and neurological disorders when considering the full transcriptome. The inclusion in our analyses of genes associated with these disorders that nevertheless present odd or low variability and asynchronous ON/OFF states proved essential to making this differentiation. Considering only HEV genes would have missed this dichotomy. Furthermore, these distinctions extended to human tissues commonly studied in genomics. It is our hope that the multilayered, more inclusive approach offered in this work paves the way to open new lines of inquiry and help advance basic research in mental and physical health in general.

## DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: <https://doi.org/10.5281/zenodo.6299834>.

## AUTHOR CONTRIBUTIONS

TB analyzed data, designed analyses, and wrote manuscript. SS curated data and edited manuscript. FHG supervised work and edited manuscript. TS supervised work and provided computational resources. ET supervised work, conceived work, analyzed data, and wrote manuscript. All authors contributed to the final state of the manuscript and agreed to its publication.

## FUNDING

This work was supported by The New Jersey Governor's Council for the Research and Treatments of Autism Grants CAUT17BSP024, CAUT18ACE014.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnins.2022.884707/full#supplementary-material>

## REFERENCES

- Ashwood, P., and Van de Water, J. (2004). Is autism an autoimmune disease? *Autoimmun. Rev.* 3, 557–562. doi: 10.1016/j.autrev.2004.07.036
- Ashwood, P., Willis, S., and Van de Water, J. (2006). The immune response in autism: a new frontier for autism research. *J. Leukoc. Biol.* 80, 1–15. doi: 10.1189/jlb.1205707
- Bell, A. J., and Sejnowski, T. J. (1996). Learning the higher-order structure of a natural sound. *Network* 7, 261–267. doi: 10.1088/0954-898X/7/2/005
- Bell, A. J., and Sejnowski, T. J. (1997). The “independent components” of natural scenes are edge filters. *Vis. Res.* 37, 3327–3338. doi: 10.1016/s0042-6989(97)00121-1
- Bermperidis, T., Rai, R., Ryu, J., Zanutto, D., Agrawal, S. K., Lalwani, A. K., et al. (2021). Optimal time lags from causal prediction model help stratify and forecast nervous system pathology. *Sci. Rep.* 11:20904. doi: 10.1038/s41598-021-00156-2
- Bernard, J. A., and Mittal, V. A. (2015). Updating the research domain criteria: the utility of a motor dimension. *Psychol. Med.* 45, 2685–2689. doi: 10.1017/S0033291715000872
- Brincker, M., and Torres, E. B. (2018). “Chapter 1 – Why study movement variability in autism,” in *Autism: the Movement Sensing Perspective*, eds E. B. Torres and C. Whyatt (Boca Raton, FL: CRC Press), 386.
- Brovedani, P., and Masi, G. (2000). Tourette’s syndrome in children: neurological, neuropsychological and psychiatric issues. *Panminerva Med.* 42, 141–149.
- Choi, K., and Torres, E. B. (2014). Intentional signal in prefrontal cortex generalizes across different sensory modalities. *J. Neurophysiol.* 112, 61–80. doi: 10.1152/jn.00505.2013
- Crossley, N. A., Scott, J., Ellison-Wright, I., and Mechelli, A. (2015). Neuroimaging distinction between neurological and psychiatric disorders. *Br. J. Psychiatry* 207, 429–434. doi: 10.1192/bjp.bp.114.154393
- David, A. S., and Nicholson, T. (2015). Are neurological and psychiatric disorders different? *Br. J. Psychiatry* 207, 373–374. doi: 10.1192/bjp.bp.114.158550
- Drton, M., and Maathuis, M. H. (2017). Structure learning in graphical modeling. *Annu. Rev. Stat. Appl.* 4, 365–393. doi: 10.1146/annurev-statistics-060116-053803
- Elsayed, M. (2021). *Characterization of Psychophysiological Responses to Pressure Pain*. Piscataway, NJ: Psychology, Rutgers The State University of New Jersey, School of Graduate Studies.
- Friston, K. J., Redish, A. D., and Gordon, J. A. (2017). Computational nosology and precision psychiatry. *Comput. Psychiatr.* 1, 2–23. doi: 10.1162/CPSY\_a\_00001
- Goodkind, M., Eickhoff, S. B., Oathes, D. J., Jiang, Y., Chang, A., Jones-Hagata, L. B., et al. (2015). Identification of a common neurobiological substrate for mental illness. *JAMA Psychiatry* 72, 305–315. doi: 10.1001/jamapsychiatry.2014.2206
- Gretton, A., Fukumizu, K., Hui Teo, C., Song, J., Scholkopf, B., and Smola, A. J. (2007). “A kernel statistical test of independence,” in *Advances in Neural Information Processing Systems 20 (NeurIPS 2007)*, Vancouver, BC.
- Gretton, A., and Györfi, L. (2010). Consistent nonparametric tests of independence. *J. Mach. Learn. Res.* 11, 1391–1423.
- GTEX Consortium (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* 369, 1318–1330. doi: 10.1126/science.aaz1776
- Hoëffding, W. (1948). A non-parametric test of independence. *Ann. Math. Stat.* 19, 293–325.
- Huys, Q. J., Maia, T. V., and Frank, M. J. (2016). Computational psychiatry as a bridge from neuroscience to clinical applications. *Nat. Neurosci.* 19, 404–413. doi: 10.1038/nn.4238
- Jekely, G., Godfrey-Smith, P., and Keijzer, F. (2021). Reafference and the origin of the self in early nervous system evolution. *Philos. Trans. R Soc. Lond. B Biol. Sci.* 376:20190764. doi: 10.1098/rstb.2019.0764
- Kawato, M., and Wolpert, D. (1998). Internal models for motor control. *Novartis Found. Symp.* 218, 291–304; discussion 304–7.
- Kereshian, J., Burd, L., and Klug, M. G. (1995). Comorbid Tourette’s disorder and bipolar disorder: an etiologic perspective. *Am. J. Psychiatry* 152, 1646–1651. doi: 10.1176/ajp.152.11.1646
- Koga, S., Murakami, A., Josephs, K. A., and Dickson, D. W. (2022). Diffuse Lewy body disease presenting as Parkinson’s disease with progressive aphasia. *Neuropathology* 42, 82–89. doi: 10.1111/ neurop.12780
- Mehta, N. (ed.) (2011). Mind-body dualism: a critique from a health perspective. *Mens Sana Monogr.* 9, 202–209. doi: 10.4103/0973-1229.77436
- Meltzer, A., and Van de Water, J. (2017). The role of the immune system in autism spectrum disorder. *Neuropsychopharmacology* 42, 284–298. doi: 10.1038/npp.2016.158
- Michel, M., Schmidt, M. J., and Mirmics, K. (2012). Immune system gene dysregulation in autism and schizophrenia. *Dev. Neurobiol.* 72, 1277–1287. doi: 10.1002/dneu.22044
- Monge, G. (1781). *Memoire sur la Theorie des Deblais et des Remblais., Histoire de l’Academie Royale des Science; avec les Memoires de Mathematique et de Physique*. Paris: De l’imprimerie Royale.
- Nguyen, J., Majmudar, U., Papatomas, T. V., Silverstein, S. M., and Torres, E. B. (2016). Schizophrenia: the micro-movements perspective. *Neuropsychologia* 85, 310–326. doi: 10.1016/j.neuropsychologia.2016.03.003
- Rogers, D. M. (1992). *Motor Disorder in Psychiatry: Towards a Neurological Psychiatry*. Chichester: Wiley & Sons.
- Rubner, Y., Tomasi, C., and Guibas, L. J. (1998). “Metric for distributions with applications to image databases,” in *Proceedings of the ICCV*, Bombay.
- Ryu, J., Bar-Shalita, T., Granovsky, Y., Weissman-Fogel, L., and Torres, E. B. (2021). Personalized biometrics of physical pain agree with psychophysics by participants with sensory over responsivity. *J. Pers. Med.* 11:93. doi: 10.3390/jpm11020093
- Ryu, J., and Torres, E. (2020). The autonomic nervous system differentiates between levels of motor intent and end effector. *J. Pers. Med.* 10:76. doi: 10.3390/jpm10030076
- Sandor, P. (1993). Gilles de la Tourette syndrome: a neuropsychiatric disorder. *J. Psychosom. Res.* 37, 211–226. doi: 10.1016/0022-3999(93)90030-j
- Slowinski, P., Alderisio, F., Zhai, C., Shen, Y., Tino, P., Bortolon, C., et al. (2017). Unravelling socio-motor biomarkers in schizophrenia. *NPJ Schizophr.* 3:8. doi: 10.1038/s41537-016-0009-x
- Torres, E. B. (2011). Two classes of movements in motor control. *Exp. Brain Res.* 215, 269–283. doi: 10.1007/s00221-011-2892-8
- Torres, E. B. (2013). Signatures of movement variability anticipate hand speed according to levels of intent. *Behav. Brain Funct.* 9:10. doi: 10.1186/1744-9081-9-10
- Torres, E. B. (2016). Rethinking the study of volition for clinical use. *Adv. Exp. Med. Biol.* 957, 229–254. doi: 10.1007/978-3-319-47313-0\_13
- Torres, E. B. (2017). “System and method for determining amount of volition in a subject,” in *Google Patents, Rutgers*, ed. USPTO (New Brunswick, NJ: The State University of New Jersey), 15.
- Torres, E. B. (2018). “Turning the tables: autism shows the social deficit of our society,” in *Autism: The Movement Sensing Perspective*, eds E. B. Torres and C. P. Whyatt (Boca Raton, FL: CRC Press), 398.
- Torres, E. B. (2020). Reframing psychiatry for precision medicine. *J. Pers. Med.* 10:144. doi: 10.3390/jpm10040144
- Torres, E. B. (2021). Precision autism: genomic stratification of disorders making up the broad spectrum may demystify its “Epidemic Rates”. *J. Pers. Med.* 11:1119. doi: 10.3390/jpm11111119
- Torres, E. B., Caballero, C., and Mistry, S. (2020). Aging with autism departs greatly from typical aging. *Sensors* 20:572. doi: 10.3390/s20020572
- Torres, E. B., and Denisova, K. (2016). Motor noise is rich signal in autism research and pharmacological treatments. *Sci. Rep.* 6:37422. doi: 10.1038/srep37422
- Torres, E. B., Isenhower, R. W., Nguyen, J., Whyatt, C., Nurnberger, J. I., Jose, J. V., et al. (2016). Toward precision psychiatry: statistical platform for the personalized characterization of natural behaviors. *Front. Neurol.* 7:8. doi: 10.3389/fneur.2016.00008
- Torres, E. B., and Lande, B. (2015). Objective and personalized longitudinal assessment of a pregnant patient with post severe brain trauma. *Front. Hum. Neurosci.* 9:128. doi: 10.3389/fnhum.2015.00128
- Torres, E. B., Brincker, M., Isenhower, R. W., Yanovich, P., Stigler, K. A., Nurnberger, J. I., et al. (2013a). Autism: the micro-movement perspective. *Front. Integr. Neurosci.* 7:32. doi: 10.3389/fnint.2013.00032
- Torres, E. B., Quian Quiroga, R., Cui, H., and Buneo, C. A. (2013b). Neural correlates of learning and trajectory planning in the posterior parietal cortex. *Front. Integr. Neurosci.* 7:39. doi: 10.3389/fnint.2013.00039
- Tu, H., Zhang, Z. W., Qiu, L., Lin, Y., Jiang, M., Chia, S. Y., et al. (2022). Increased expression of pathological markers in Parkinson’s disease dementia

- post-mortem brains compared to dementia with Lewy bodies. *BMC Neurosci.* 23:3. doi: 10.1186/s12868-021-00687-4
- van der Maaten, L., and Hinton, G. (2008). Visualizing data using t-SNE. *J. Mach. Learn. Res.* 9, 2579–2605.
- Vaskevich, A., and Torres, E. B. (2022). Rethinking statistical learning as a dynamic stochastic process, from the motor systems perspective. *bioRxiv* [Preprint]. doi: 10.1101/2022.01.18.476831
- Von Holst, E., and Mittelstaedt, H. (1950). “The principle of reafference: interactions between the central nervous system and the peripheral organs,” in *Perceptual Processing: Stimulus Equivalence and Pattern Recognition*, ed. P. C. Dodwell (New York, NY: Appleton-Century-Crofts), 41–72.
- Walther, S., and Mittal, V. A. (2017). Motor system pathology in psychosis. *Curr. Psychiatry Rep.* 19:97. doi: 10.1007/s11920-017-0856-9
- Walther, S., and Strik, W. (2012). Motor symptoms and schizophrenia. *Neuropsychobiology* 66, 77–92. doi: 10.1159/000339456
- Walther, S., van Harten, P. N., Waddington, J. L., Cuesta, M. J., Peralta, V., Dupin, L., et al. (2020). Movement disorder and sensorimotor abnormalities in schizophrenia and other psychoses – European consensus on assessment and perspectives. *Eur. Neuropsychopharmacol.* 38, 25–39. doi: 10.1016/j.euroneuro.2020.07.003
- Walther, S., Vladimirova, I., Alexaki, D., Schappi, L., Damme, K. S. F., Mittal, V. A., et al. (2022). Low physical activity is associated with two hypokinetic motor abnormalities in psychosis. *J. Psychiatr. Res.* 146, 258–263. doi: 10.1016/j.jpsychires.2021.11.014
- White, P., Rickards, H., and Zeman, A. (2012). Time to end the distinction between mental and neurological illnesses. *BMJ* 344:e3454. doi: 10.1136/bmj.e3454
- Wolpert, D. M., and Kawato, M. (1998). Multiple paired forward and inverse models for motor control. *Neural Netw.* 11, 1317–1329. doi: 10.1016/s0893-6080(98)00066-5
- Wu, D., Jose, J. V., Nurnberger, J. I., and Torres, E. B. (2018). A biomarker characterizing neurodevelopment with applications in autism. *Sci. Rep.* 8:614. doi: 10.1038/s41598-017-18902-w
- Yao, Z., Mich, J. K., Ku, S., Menon, V., Krostag, A. R., Martinez, R. A., et al. (2017). A single-cell roadmap of lineage bifurcation in human ESC models of embryonic brain development. *Cell Stem Cell* 20, 120–134. doi: 10.1016/j.stem.2016.09.011
- Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Publisher’s Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.
- Copyright © 2022 Bermperidis, Schafer, Gage, Sejnowski and Torres. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.