

# Investigating Dopaminergic Abnormalities in Psychosis with Normative Modelling and Multisite Molecular Neuroimaging

A. Giacomel<sup>1\*</sup>, D. Martins<sup>1,2</sup>, G. Nordio<sup>1</sup>, R. Easmin<sup>1</sup>, O. Howes<sup>3</sup>, Pierluigi Selvaggi<sup>1,4</sup>, S.C.R. Williams<sup>1</sup>, F. Turkheimer<sup>1</sup>, M. De Groot<sup>5</sup>, O. Dipasquale<sup>1\*\*</sup>, M. Veronese<sup>1,6\*\*</sup> and the FDOPA PET imaging working group

<sup>1</sup>Centre for Neuroimaging Sciences, Institute of Psychology, Psychiatry and Neuroscience (IoPPN), King's College London, London, UK

<sup>2</sup>Division of Adult Psychiatry, Department of Psychiatry, Geneva University Hospitals, Rue Gabrielle Perret-Gentil 4, 1205, Geneva, Switzerland

<sup>3</sup>Department of Psychosis Studies, Institute of Psychology, Psychiatry and Neuroscience (IoPPN), King's College London, London, UK

<sup>4</sup>Department of Translational Biomedicine and Neuroscience, University of Bari "Aldo Moro", Bari, Italy

<sup>5</sup>GSK R&D, Clinical Pharmacology and Experimental Medicine, Clinical Imaging, Stevenage, UK

<sup>6</sup>Department of Information Engineering, University of Padova, Padova, Italy

\*Corresponding authors: [alessio.giacomel@kcl.ac.uk](mailto:alessio.giacomel@kcl.ac.uk); [mattia.veronese@unipd.it](mailto:mattia.veronese@unipd.it)

\*\*Shared senior authorship

FDOPA PET imaging working group (in alphabetical order):

Ilinca Angelescu<sup>1,2</sup>, Michael Bloomfield<sup>3</sup>, Ilaria Bonoldi<sup>1</sup>, Faith Borgan<sup>1</sup>, Enrico D'Ambrosio<sup>1,4</sup>, Tarik Dahoun<sup>5,6,7</sup>, Arsime Demjaha<sup>1</sup>, Alice Egerton<sup>1</sup>, Sameer Jauhar<sup>1,5,6</sup>, Stephen Kaar<sup>1,5,8,9</sup>, Euitae Kim<sup>10,11,12</sup>, Seoyoung Kim<sup>10</sup>, James MacCabe<sup>1</sup>, Julian Matthews<sup>13</sup>, Robert McCutcheon<sup>1,14,15</sup>, Philip McGuire<sup>14</sup>, Chiara Nosarti<sup>16,17</sup>, Maria Rogdaki<sup>1</sup>, Grazia Rutigliano<sup>1,5</sup>, Peter S Talbot<sup>18</sup>, Luke Vano<sup>1,5</sup>

<sup>1</sup> Department of Psychosis Studies, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, United Kingdom

<sup>2</sup> Max Planck UCL Centre for Computational Psychiatry and Ageing Research, Institute of Neurology, London, UK

**NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.**

<sup>3</sup> Division of Psychiatry, Faculty of Brain Sciences, University College of London, London, UK

<sup>4</sup> Group of Psychiatric Neuroscience, Department of Translational Biomedicine and Neuroscience, University of Bari "Aldo Moro", Bari, Italy

<sup>5</sup> Psychiatric Imaging Group, MRC London Institute of Medical Sciences, Hammersmith Hospital, Imperial College London, London, UK

<sup>6</sup> Institute of Clinical Sciences, Faculty of Medicine, Imperial College, Imperial College London, London, UK

<sup>7</sup> Department of Psychiatry, Warneford Hospital, University of Oxford, Oxford, UK

<sup>8</sup> Institute of Clinical Sciences (ICS), Faculty of Medicine, Imperial College London, UK

<sup>9</sup> South London and Maudsley NHS Foundation Trust, London, UK

<sup>10</sup> Department of Psychiatry, Seoul National University Bundang Hospital, Gyeonggi-do, Republic of Korea

<sup>11</sup> Department of Psychiatry, College of Medicine, Seoul National University, Seoul, Republic of Korea <sup>12</sup>

Department of Brain & Cognitive Sciences, College of Natural Sciences, Seoul National University, Seoul, Republic of Korea

<sup>13</sup> Division of Neuroscience and Experimental Psychology, School of Biological Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester, UK

<sup>14</sup> Department of Psychiatry, Warneford Hospital, University of Oxford, Oxford, OX3 7JX, UK

<sup>15</sup> Oxford Health NHS Foundation Trust, Oxford, OX3 7JX, UK

<sup>16</sup> Department of Child and Adolescent Psychiatry, Institute of Psychiatry, Psychology & Neurosciences, King's College London, London, UK

<sup>17</sup> Centre for the Developing Brain, Division of Imaging Sciences & Biomedical Engineering, King's College London, London, UK

<sup>18</sup> Division of Neuroscience and Experimental Psychology, School of Biological Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester, UK

## Abstract

Molecular neuroimaging techniques, like PET and SPECT, offer invaluable insights into the brain's in-vivo biology and its dysfunction in neuropsychiatric patients. However, the transition of molecular neuroimaging into diagnostics and precision medicine has been limited to a few clinical applications, hindered by issues like practical feasibility and high costs. In this study, we explore the use of normative modelling (NM) for molecular neuroimaging to identify individual patient deviations from a reference cohort of subjects. NM potentially addresses challenges such as small sample sizes and diverse acquisition protocols that are typical of molecular neuroimaging studies. We applied NM to two PET radiotracers targeting the dopaminergic system ( $[^{11}\text{C}]$ -(+)-PHNO and  $[^{18}\text{F}]$ FDOPA) to create a normative model to reference groups of controls. The models were subsequently utilized on various independent cohorts of patients experiencing psychosis. These cohorts were characterized by differing disease stages, treatment responses, and the presence or absence of matched controls. Our results showed that patients exhibited a higher degree of extreme deviations (~3-fold increase) than controls, although this pattern was heterogeneous, with minimal overlap in extreme deviations topology (max 20%). We also confirmed the value of striatal  $[^{18}\text{F}]$ FDOPA signal to predict treatment response (striatal AUC ROC: 0.77-0.83). Methodologically, we highlighted the importance of data harmonization before data aggregation. In conclusion, normative modelling can be effectively applied to molecular neuroimaging after proper harmonization, enabling insights into disease mechanisms and advancing precision medicine. The method is valuable in understanding the heterogeneity of patient populations and can contribute to maximising cost efficiency in studies aimed at comparing cases and controls.

**Keywords:** Normative Modelling, Positron Emission Tomography, Dopamine, Psychosis

**Running title:** Normative Modelling in Molecular Neuroimaging

## Introduction (1500 words)

Psychosis is a debilitating mental health condition characterised by a loss of contact with reality, abnormal thoughts, perceptions, and behaviour. While psychosis is a hallmark of schizophrenia, psychotic symptoms can also be present in other psychiatric disorders including bipolar disorder, and severe depression. Psychosis has been the subject of extensive neuroimaging research aimed at unravelling its underlying neurobiological mechanisms. Consistent findings from structural magnetic resonance imaging (MRI) studies have systematically shown brain structure abnormalities, such as reduced grey matter volume and compromised white matter integrity<sup>1</sup>. Similarly, functional MRI (fMRI) studies have identified abnormal connectivity and atypical brain activation patterns during cognitive tasks, suggesting dysfunction in cortical and subcortical regions associated with attention, memory, and emotion regulation<sup>2</sup>. Molecular neuroimaging, such as Positron Emission Tomography (PET), has also provided valuable insights into the molecular alterations associated with psychosis, highlighting the involvement of multiple neuroreceptor systems, metabolism<sup>3</sup> and neuroinflammation<sup>4</sup>.

Abnormalities in dopamine neurotransmission have emerged as a consistent finding in psychosis<sup>5</sup>. The dopamine hypothesis posits dysregulation within the dopaminergic system as a key factor in the development of psychosis<sup>6</sup>. PET studies have consistently shown elevated dopamine synthesis capacity in specific brain regions, particularly the striatum, in individuals with psychosis<sup>7</sup>. Specifically, this hyperactivity of the striatal dopaminergic system is believed to contribute to positive symptoms such as hallucinations and delusions<sup>8-10</sup>. The glutamate hypothesis has also gained significant attention in psychosis research. This hypothesis suggests dysfunction in the glutamatergic system, particularly N-methyl-D-aspartate (NMDA) receptor hypofunction, as a contributing factor for psychosis<sup>11,12</sup>. [<sup>18</sup>F]FDOPA PET studies have also provided evidence of altered dopamine-glutamate interactions, highlighting the intricate relationship between these neurotransmitter systems<sup>5</sup>.

While offering unique insights into the brain mechanisms underlying psychosis, most of the neuroimaging literature has primarily focused on attempting to identify single unifying

pathophysiological processes shared across patients. This has been achieved using 'standard' cross-sectional statistics based on group averages, often treating individual differences merely as noise. However, there is a growing awareness that these conventional approaches fall short of fully capturing the multifaceted characteristics of complex mental health disorders like psychosis. To develop imaging-based biomarkers with true clinical utility, interindividual variability cannot be simply dismissed as noise or assumed to be part of measurement variability. This awareness has been fostered, in part, by the parallel tendency of collecting large archives of neuroimaging data, which has triggered a gradual shift in neuroimaging towards employing advanced analytical modelling methods to model subjects' characteristics<sup>13</sup>. By providing statistical inferences at the individual level with respect to an expected pattern, these methods offer the opportunity to parse heterogeneity across cohorts and identify the unique characteristics of each individual. Furthermore, by focusing on personalized perspective and acknowledging the potential to serve as fundamental keys in the creation of neuroimaging biomarkers for furthering understanding of the neurobiological basis of psychiatric disorders and potentially predicting treatment outcomes<sup>14,15</sup>.

One of these advanced modelling methods that has gained widespread traction in neuroimaging research is normative modelling (NM). This statistical framework is based on the concept of paediatric growth charts<sup>16</sup>, which utilise a series of percentile curves to illustrate the normal distribution of children's body measurements, such as weight, height, and head circumference, as a function of their age. When applied to neuroimaging data<sup>14,17,18</sup>, this approach enables the identification of a relationship between quantitative neuroimaging biomarkers, such as regional brain volume or thickness measured with MRI, and relevant factors like specific clinical, demographic, or behavioural measures of interest<sup>17</sup>.

The rationale for employing normative modelling in neuroimaging is twofold. Firstly, this approach allows us to use neuroimaging data from healthy individuals to establish the normal range for a specific brain characteristic. By describing a portion of the between-subject variability of the given neuroimaging measure through demographic factors (e.g., age, sex<sup>19,20</sup>, BMI) or abilities (e.g., IQ<sup>21</sup>), normative modelling defines what can be

considered typical or within the range of expected variation. Secondly, it allows for the quantification of regional deviations from normality at the individual level. This aspect aims to identify disease-specific patterns of alterations and dissect disease heterogeneity across patients. By comparing an individual's neuroimaging data to the established normal range, it is possible to pinpoint aberrations that may indicate the presence of certain conditions and contribute significantly to the understanding and characterization of brain disorders<sup>14,22</sup>.

Up to now, normative modelling has found extensive application in the identification of consistent disease-specific patterns of brain structural alterations, as demonstrated in schizophrenia<sup>19,20,23</sup>, bipolar disorder<sup>19,20</sup>, ADHD<sup>24</sup>, and Parkinson's disease<sup>25</sup>, using structural MRI data (i.e., T1-weighted MRI<sup>19</sup> and diffusion-weighted imaging<sup>26</sup>). However, this methodology has yet to be applied to the molecular underpinnings of human brain function. The primary challenge in applying normative modelling to PET/SPECT brain imaging lies in the necessity of pooling large datasets to establish reliable parameters for the normative models. To estimate normative models of MRI-based brain measures, hundreds<sup>19,24,27</sup> or even thousands<sup>28</sup> of scans are typically included in the reference cohort. However, these numbers cannot be feasibly achieved in single molecular neuroimaging studies due to the substantial costs of PET and SPECT scans (up to 10 times higher than MRI scans) and ethical issues related to the use of radioactive tracers for research purposes. Nevertheless, recent developments in the molecular neuroimaging community, including a greater willingness to share data<sup>29,30</sup> and the establishment of international consortia (e.g., ENIGMA<sup>31</sup>), along with the development of effective harmonisation techniques for neuroimaging data<sup>32-35</sup>, have paved the way for the use of normative modelling in molecular neuroimaging.

In this work, our primary objective is to demonstrate the feasibility of employing normative modelling to molecular neuroimaging in the context of psychosis studies. Despite the challenges posed by small sample sizes and diverse experimental designs, we hypothesise that normative modelling can be effectively applied to this type of data, providing valuable insights into molecular brain functions in healthy individuals and those experiencing psychosis. Furthermore, we hypothesise that normative modelling can serve

as a powerful tool to identify both the magnitude and the spatial distribution of molecular alterations at the individual patient level by 1) extending analysis of dopamine alterations in the striatum to the whole brain, 2) identifying common patterns of dopamine dysfunction across multiple and independent datasets of patients with psychosis, and 3) linking inter-individual deviation from normalities to clinical symptoms and response to treatment.

This study is hence organised into two main parts. In the first section, we used two datasets<sup>36,37</sup> of [<sup>11</sup>C]-(+)-PHNO PET imaging data measuring D<sub>2/3</sub> dopamine receptor density, acquired with two scanners, to compare the effects of two image harmonisation methods on the distribution of the deviation scores of the NM and evaluated which one is the most effective at reducing the scanner effects. We then applied the so-identified optimal harmonisation method to studies of dopamine synthesis capacity using [<sup>18</sup>F]FDOPA PET imaging in healthy controls (HC), acquired with five different scanners, and assessed if the [<sup>11</sup>C]-(+)-PHNO results were generalizable to a different radiotracer and a greater number of scanners.

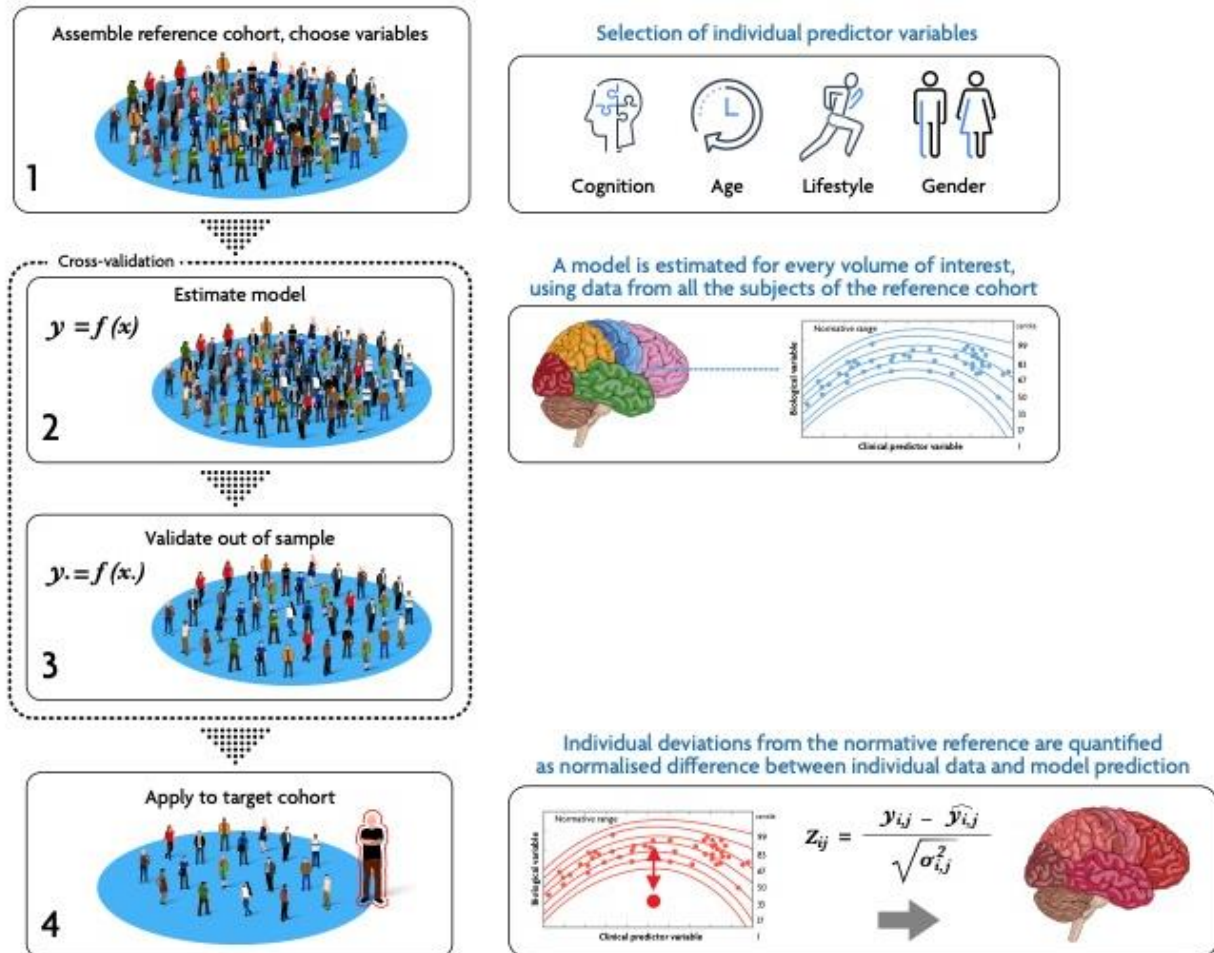
In the second section, we used the estimated [<sup>18</sup>F]FDOPA PET model to calculate the deviations from normality in four datasets<sup>38-47</sup> acquired in patients with psychosis, one of which did not include any matched HC. Here, we investigated the presence of shared spatial patterns of deviation in psychosis, but also differences between patients with first-episode psychosis and chronic schizophrenia. In addition, we tested the feasibility of using independent datasets of patients with no matched HCs when a big enough reference cohort is already available. Here, we assessed the replicability of the findings from the previous comparisons and investigated the clinical value of extreme deviations by looking at relationships between patient-specific extreme deviations from normality and clinical symptoms. Lastly, since [<sup>18</sup>F]FDOPA PET imaging has been recently proposed as a potential biomarker for treatment stratification in psychosis<sup>38</sup>, we built a classifier to assess the value of [<sup>18</sup>F]FDOPA PET NM for predicting treatment response, and compared its performances to the reference standard analytics, to evaluate if these would be comparable or outperforming.

## Methods

### Box 1: Normative Modelling Theory

The general framework to derive normative models from neuroimaging data is explained in detail elsewhere<sup>14,17,22</sup>. In brief, it comprises several steps (Figure B1). Firstly, a reference cohort of healthy controls (HC) is chosen. Next, a specific brain measure, such as a given summary measure at the whole-brain level, region-of-interest (ROI)-based derivatives or voxel-wise brain data from structural MRI scans (e.g., volume and cortical thickness), is selected. Additionally, a set of variables (i.e., predictors) is chosen to explain the brain measure. Statistical models are then constructed for each summary measure, establishing a connection between the neuroimaging data and the selected predictors. The primary outcome of the NM is a measure of deviation, typically expressed as a z-score. This z-score indicates the extent to which the specific brain measure deviates from the normative distribution, providing valuable information about its deviation from the reference group. Following the model's estimation, its ability to link individual covariates to neuroimaging data needs to be further evaluated both in-sample (e.g., using k-fold cross-validation) and out-of-sample (e.g., using an independent cohort of HC)<sup>22</sup>. Once the model is validated, it can be used to estimate the deviations of a target cohort (generally corresponding to patients), which can then be analysed by investigating the magnitude and spatial pattern of the so-called “extreme deviations”. These extreme values represent brain areas deviating more than two standard deviations from the reference mean.





Steps of normative modelling. Main steps for the generation and application of normative modelling to neuroimaging data. In brief, the main steps include (1) the assembly of a reference cohort (usually HC) and the selection of covariates of interest (e.g., demographics or cognitive). (2) Model estimation, at desired level of granularity (i.e., voxel- or ROI-level). (3) Model validation, if possible, out-of-sample, or using cross-validation. (4) Application to target cohort, usually patients, and estimation of deviation scores (Z-scores).

## Study datasets

### [<sup>11</sup>C]-(+)-PHNO Datasets

Data from 77 HC (Table 1) from two previous studies<sup>36,37</sup>, acquired with [<sup>11</sup>C]-(+)-PHNO, were used to assess and select the best image harmonisation method. The scans were acquired with two different scanners (PET/CT, Siemens Hi-Rez Biograph 6, N=54; and PET/MR, Siemens Biograph mMR, N=23) in the same imaging site (Invicro, London). Experimental designs and imaging protocols were consistent between the two studies. Further details on data acquisition, image processing, and data analysis are reported in the original publications<sup>36,37</sup>. In short, dopamine D2/3 receptor density was measured using non-displaceable binding potential ( $BP_{ND}$ <sup>48</sup>) as the parameter of interest. For both scanners, parametric  $BP_{ND}$  images were obtained using the same MATLAB-based pipeline with a simplified reference tissue model with cerebellar grey matter used as the reference region. Final maps were all normalised to MNI152 standard space before image harmonisation and NM.

### [<sup>18</sup>F]FDOPA Datasets

6-[<sup>18</sup>F]-fluoro-L-DOPA (FDOPA) data from 142 HC and 109 patients with psychosis from previous studies<sup>38-47</sup> were used for estimating the NM of dopamine synthesis capacity and exploration of its alterations in psychosis (Table 1). The datasets consisted of two datasets of HC only (FDOPA\_HC01 and FDOPA\_HC02), three case-control datasets (FDOPA\_01, FDOPA\_02, and FDOPA\_03) and one dataset of patients only (FDOPA\_04). Data were acquired with five different scanners (Siemens Hi-Rez Biograph 6, Siemens Biograph 40 TruePoint, Siemens TruePoint 6, ECAT HR+ 962, and ECAT EXACT 3D) in three different imaging sites (MRC Cyclotron Unit, London; Invicro, London and Bundang Hospital, South Korea). The acquisition protocol was consistent across sites. All FDOPA PET imaging sessions were acquired with a continuous dynamic acquisition (no blood sampling), with scanning beginning with the tracer injection and lasting for ~90 minutes. All participants received carbidopa (150 mg) and entacapone (400 mg) orally 1 hour before imaging to increase the brain tracer uptake and reduce the peripheral formation of radiolabelled metabolite, respectively. The FDOPA tracer (injected dose ranging from 86.4 to 414.4 MBq,) was administered by intravenous bolus injection after the acquisition of a brain CT or MRI for attenuation correction, depending on the scanner availability at each imaging site. PET data reconstruction varied across imaging sites and scanner

types, but all included correction for random noise, scatter, and tissue attenuation. Dopamine synthesis capacity indicated by the parameter  $K_i^{cer}$  (sometimes indicated as  $K_i$ ,  $\text{min}^{-1}$ ) was quantified with the Gjedde-Patlak graphical method using the cerebellum as the reference region. Data were normalised to MNI152 standard space using the same in-house pipeline<sup>49</sup>, removing data with poor quality and excess of motion.

Patients consisted of individuals experiencing either a first episode of psychosis (FEP) or diagnosed with a chronic psychotic disorder (Table 1). In brief, FDOPA\_01 patients were recruited if they met structured clinical interview for DSM-IV axis I disorder criteria for schizophrenia (mean Total PANSS [SD]: 72.94 [16.46]), had at least one current or previous psychotic episode (duration of illness, median [IQR]: 24 [24] months), and were antipsychotic medication naïve (no current or previous treatment) or free (not taking medication at scanning time with at least 6 months washout for oral medication or 6 months for depot medication)<sup>39</sup>. FDOPA\_02 patients were composed of patients with chronic illness (duration of illness, mean [SD]: 128 [112.24] months) and were recruited if they met the following criteria: they met the DSM-IV criteria for schizophrenia, had a total score  $\leq 80$  in the PANSS Total scale (mean Total PANSS [SD]: 50.0 [9.64]), had received first-line antipsychotic drugs (including risperidone, olanzapine, and paliperidone) or clozapine for at least 12 weeks<sup>43</sup>. Patients in the FDOPA\_03 cohort were also patients with chronic illness who met the DSM-IV criteria for schizophrenia (duration of illness, mean [SD]: 193.8 [367.08] months), received at least two sequential antipsychotic trials (of at least 4 weeks duration), had a Total PANSS score  $\geq 75$ , and were not taking clozapine at time of scanning<sup>8</sup>. Lastly, FDOPA\_04 patients were included if they met the DSM-V criteria for schizophrenia or schizophreniform disorders, and were not medicated with clozapine in the previous 3 months to scanning<sup>44</sup>. Full information on the recruitment criteria and clinical characteristics of the patient population is reported in the original papers<sup>38-47</sup>.

For a subset of patients of the cohorts FDOPA\_01 and FDOPA\_04, symptom severity measures (i.e., Positive and Negative Symptoms Scale – PANSS<sup>50</sup>) were also available.

## Data harmonisation strategies

To harmonise the [<sup>11</sup>C]-(+)-PHNO scans we employed two methods widely used in the literature: Gaussian kernel smoothing<sup>34</sup> and Combat harmonisation<sup>32,33</sup>.

**Gaussian kernel smoothing** (hereafter referred to as “smoothing”): This method involves applying a 3D convolution filter with a Full Width Half Maximum (FWHM) of  $2.35\sigma$ , where  $\sigma$  represents the kernel size. For this step, we used FSL maths<sup>51-53</sup> setting a kernel size of 3mm, determined through iterative analysis to achieve the best match between the reference and the target data to be harmonised (i.e., PET/CT and PET/MR, respectively).

**Combat harmonisation**: Combat is a Bayesian harmonisation method derived from the field of genomics<sup>54</sup>, which assumes site-specific scale- and shift-factors to be estimated between different batches (e.g., scanners or sites). Unlike smoothing, Combat allows to estimate and preserve the effect of certain covariates defined a priori on the data variability. In our specific case, we employed the NeuroCombat python library (v.0.2.10+)<sup>32,33</sup> and preselected the same covariates subsequently used for NM (i.e., age, sex for both tracers and BMI for [<sup>11</sup>C]-(+)-PHNO).

To assess the best harmonisation method, we evaluated the effects of harmonisation on the NM results, particularly focusing on the Z-score distributions and the variance explained by the model (see section ‘Comparison of the harmonisation methods’ below).

## **NM implementation**

After data harmonisation, Bayesian Linear Regression (BLR) was used to estimate voxel-wise NMs (i.e., one model per voxel) of dopamine receptor density and dopamine synthesis capacity from [<sup>11</sup>C]-(+)-PHNO and [<sup>18</sup>F]FDOPA data of the HC samples. BLR is a probabilistic approach to linearly model the relationship between a dependent variable and a set of independent variables. In this case, the selected independent variables were sex, age, and BMI for [<sup>11</sup>C]-(+)-PHNO, and age and sex for [<sup>18</sup>F]FDOPA. Of note, these

covariates were used in previous in-house analyses after assessing them to be significant for the specific tracers<sup>49,55–57</sup>.

NMs were estimated using the PCNToolkit (v.1.20.0) Python library, employing default non-informative priors for the parameters and the Powell optimizer for faster convergence of model fitting<sup>14</sup>. Consistently with the PET signal distribution, NM was restricted to the grey matter voxels by thresholding and binarizing a standard probabilistic map of the grey matter at 30% probability. Thresholding for GM is not typically applied for FDOPA brain PET imaging<sup>49</sup>, but this threshold was used as a trade-off between brain coverage and the computational cost of NM quantification. As a result, only voxels with at least 30% probability of being grey matter were included in the mask, yielding a total of 199,715 voxels (and hence models) for each tracer. Importantly, for the [<sup>11</sup>C]-(+)-PHNO data we estimated three different NMs, i.e., one for the unharmonized data and one for each harmonisation strategy (smoothing and Combat) to identify the strategy that best minimises residual scanner-related differences for PET data. Based on the outcome of this analysis of performance, we then used the identified method to harmonise the [<sup>18</sup>F]FDOPA datasets before the NM step.

The estimated NMs were then used to estimate the voxel-wise maps of deviation from normality for each individual. These maps, expressed in terms of Z-scores, measure the distance of a given data point in relation to the average and standard deviation of the posterior probability, weighted for the subject-specific covariates. These values were estimated using k-fold cross-validation (k=5) for the HC, or the full [<sup>18</sup>F]FDOPA NM for patients with psychosis. Furthermore, we identified voxels with an extreme deviation from normality (i.e., voxels with PET signal intensity significantly different from the reference distribution). Such extreme deviations were defined using a thresholding approach:  $Z > 2$  for *extreme positive deviations*,  $Z < -2$  for *extreme negative deviations*, and  $|Z| > 2$  for *total extreme deviations*. As the deviation scores follow a normal distribution, it is essential to note that there is an expected percentage of extremely deviating voxels. This consists of 2.5% residual density on each tail of the distribution (i.e.,  $Z > 2$  or  $Z < -2$ ), totalling 5% for the absolute value (i.e.,  $|Z| > 2$ ). Conversely, we anticipate a higher percentage of extremely deviating voxels, either positive or negative, in subjects deviating from normality.

## **Identification of the optimal harmonisation strategy based on [11C]-(+)-PHNO data**

For each harmonisation strategy (no harmonisation, harmonisation with smoothing and harmonisation with combat), we compared the average and standard deviation of the individual Z-score distributions between the two scanners (i.e., PET/CT and PET/MR) to identify any residual differences between datasets. These comparisons were performed using Welch's two-sample t-tests, implemented with the `rstatix` package<sup>58</sup> in R (version 4.2.1). Furthermore, using the brain parcellation defined by the Hammersmith atlas<sup>59</sup>, for each harmonisation strategy we estimated the regional percentage of extreme positive and negative deviations across subjects of each dataset (i.e., PET/CT and PET/MR data). We then calculated the regional differences between the two datasets and employed Wilcoxon's one-sample rank-tests implemented with the `rstatix` package<sup>58</sup> in R 4.2.1 to identify any significant scanner-related differences.

## **Validation of the optimal harmonisation strategy on [<sup>18</sup>F]FDOPA data**

We validated the harmonization method identified as optimal for the [11C]-(+)-PHNO data on the [<sup>18</sup>F]FDOPA data, to see if it would yield consistent results in terms of removing any scanner-related differences across datasets.

First, we harmonised the [<sup>18</sup>F]FDOPA HC data with the optimal harmonisation strategy identified using the [11C]-(+)-PHNO data. Then, in order to include in the analysis only those voxels for which the NMs were able to describe the between-subject variability based on the variables used to estimate the model (i.e., age and sex for the [<sup>18</sup>F]FDOPA data), we discarded the grey matter voxels where NMs were not able to converge to a solution, i.e., those reporting an explained variance (EXPV) lower than 0. Furthermore, since a positive yet very small EXPV does not necessarily mean that the model is able to meaningfully describe the data, we decided to restrict the subsequent analyses to voxels with  $EXPV > 3\%$  (i.e., voxels where NM reached statistical significance of  $p_{uncorr} < 0.05$ ). For completeness, we repeated the analyses on the data including voxels with  $EXPV > 0$ ,



and using a conservative threshold of  $EXPV \geq 10\%$ , corresponding to the optimal point from the L-curve distribution of EXPV.

Finally, to validate the optimal harmonisation strategy on [ $^{18}\text{F}$ ]FDOPA HC data, we calculated the mean and standard deviation of the individual distributions of HC Z-score deviations and compared them by means of one-way ANOVAs using the `rstatix`<sup>58</sup> library in R 4.2.1.

## **NM of [ $^{18}\text{F}$ ]FDOPA PET in psychosis**

After calculating the z-score deviations of the three clinical cohorts using the NMs estimated from the HC data, we grouped the Z-score distributions for HC and patients separately and compared them. We also calculated the Risk-Ratios (RR)<sup>60</sup> by counting the number of voxels whose Z-scores showed extreme deviations from normality, using the `epitools` library in R 4.2.1. In this case, the RR indicates the likelihood of a new individual of belonging to the clinical cohort as compared to the group of HCs.

We then tested the hypothesis that patients would show a greater overlap of extreme deviations than HC, indicating spatial consistency in the manifestation of the disorder, by investigating the spatial patterns of extreme deviations in HC and patients. This was done by counting, for each voxel and group, the number of subjects showing an extreme deviation in that voxel. This was assessed both on the positive and negative extreme deviations separately and on the total extreme deviations.

We then analyzed the differences between HC and patients in terms of the magnitude of deviation from normality by performing both a voxel-wise analysis on the thresholded Z-score maps, and on summary measures of deviations. The voxel-wise analysis allows spatial localisation of the significant differences between the two groups, whereas summary measures determine whether it is possible to collapse meaningful information into a few scores per subject which might be useful for clinical applications. The four summary scores analysed were the mean z-score, corresponding to the mean

z-score across voxels for each subject, and three measures of extreme deviations from normality, estimated by counting the percentage of voxels with Z-score  $> 2$  (i.e., positive deviations), Z-score  $< -2$  (i.e., negative deviations) or  $|Z\text{-score}| > 2$  (i.e., total deviations). Since the clinical datasets included subjects with FEP and chronic psychosis, we ran two-way ANOVAs testing the main effects of cohort (HC, FEP and chronic psychosis) and dataset (FDOPA\_01, FDOPA\_02, FDOPA\_03), and their interaction. The voxel-wise tests were performed using FSL randomise<sup>61</sup>, with 5,000 permutations and considering a cluster significant if  $p_{\text{FWE}} < 0.05^{62,63}$ , corrected for multiple comparisons using the threshold-free cluster enhancement<sup>62</sup> (TFCE) option. For the significant contrasts, we extracted the mean z-score values from the significant clusters and performed post-hoc tests using the rstatix<sup>58</sup> library in R 4.2.1. For the tests on the summary metrics, we used the rstatix<sup>58</sup> library in R 4.2.1.

Lastly, we investigated which biological pathways were driving the spatial pattern of extreme deviations identified in the voxel-wise two-way ANOVA. This was done by running imaging-transcriptomics with the imaging-transcriptomics toolbox<sup>64-66</sup> on the F-stat map reporting the main effect of the group, and by running GSEA on molecular function, as defined by the GO Molecular Function<sup>67,68</sup>.

## **Extending normative model validity to an independent patient cohort**

To demonstrate the value of NM analysis when using a clinical cohort without its own group of matched HCs and test the replicability of the results from the previous analyses, we applied the [<sup>18</sup>F]FDOPA PET NM on an independent dataset of patients with psychosis (FDOPA\_04). Of note, although this clinical cohort was independent, its scanner effect was modelled in the estimated NM since this dataset was acquired at the same site and using the same scanner as the FDOPA\_01 dataset. We estimated the four summary measures (mean Z-score and positive, negative, and total extreme deviations) in this cohort and compared them with the summary measures of the HC of the other datasets (FDOPA 01, 02 and 03). For this analysis, we performed Welch two-sample t-tests using the rstatix<sup>58</sup> library (R 4.2.1).



## Relationships between deviation scores and clinical symptoms

In datasets where clinical data were available, specifically FDOPA\_01 (comprising FEP patients) and FDOPA\_04 (comprising a combination of chronic and FEP patients), we tested the hypothesis that NM deviation scores would be associated with symptom severity, as measured using PANSS scores<sup>50</sup> using voxel-wise comparisons and summary measures. Voxel-wise covariance analysis was conducted utilizing FSL *randomise*<sup>52,53,61</sup> (independent variables: PANSS scores; dependent variable: Z-score), while Spearman correlations between whole brain summary measures and PANSS scores were computed using *rstatix*<sup>58</sup> (R 4.2.1). For the cohort of patients without matched controls (FDOPA\_04), we additionally performed a correlation analysis between the individual PANSS scores and the average Z-scores of the statistically significant clusters, identified from the voxel-wise cross-sectional analysis (Spearman correlation, *rstatix*<sup>58</sup> library, R 4.2.1).

## Prediction of Treatment Response in Patients

To determine whether the predictive power of FDOPA summary measures outperformed reference analysis (i.e., striatal  $K_i$ ), in predicting antipsychotic treatment response for each clinical cohort we constructed ROC curves for all summary measures of the deviation scores (average Z-score, positive, negative, and total extreme deviations) across the whole brain and striatum and ROC curves for the original striatal  $K_i$  measures. ROC curves were compared using the DeLong test. All analyses were performed using the *pROC* library<sup>69</sup> R 4.2.1.

## Results

### Estimation of NM for [<sup>11</sup>C]-(+)-PHNO PET and assessment of the optimal harmonisation strategy

As shown in Figure 1A, the Z-score distributions of the non-harmonised [<sup>11</sup>C]-(+)-PHNO PET data present two distinct scanner-related patterns, which are attenuated after harmonisation with both harmonisation strategies. When comparing the averages of the distributions between scanner types, we found statistically significant differences between the PET/CT and PET/MR scanners both in the unharmonized data ( $t=2.46$ ,  $p<0.01$ ) and in the data harmonised with smoothing ( $t=2.85$ ,  $p<0.01$ ), while there were no scanner-related differences in the data harmonised with Combat ( $t=0.71$ ,  $p=n.s.$ ). In terms of standard deviation, we found significant differences between scanner types in the non-harmonised data ( $t=-14.59$ ,  $p<0.001$ ) and residual differences in the data harmonised with Combat ( $t=-2.47$ ,  $p<0.01$ ), while no differences between scanner types were found in the data harmonised with smoothing ( $t=-0.40$ ,  $p=n.s.$ ).

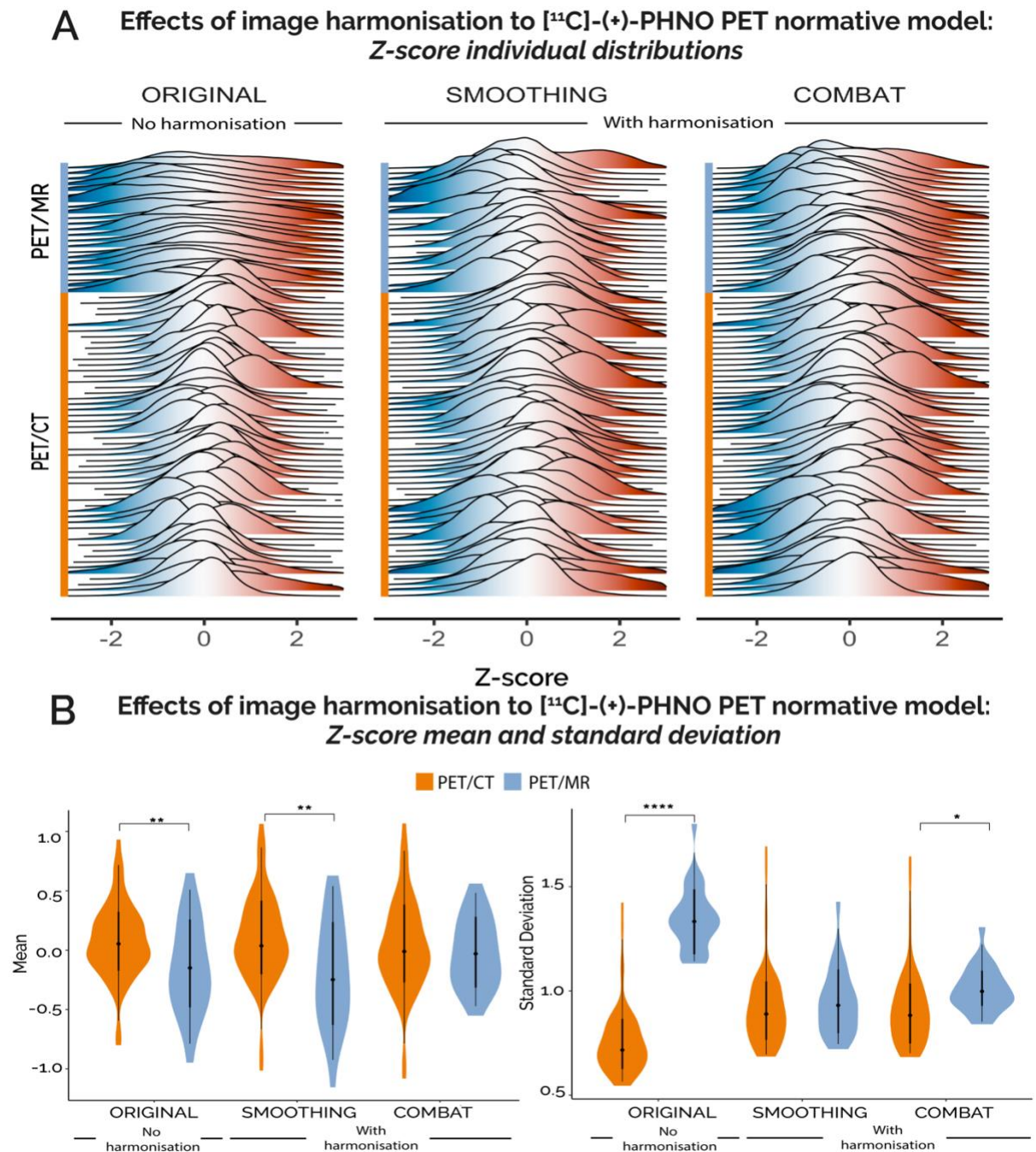


Figure 1. Harmonisation results of [ $^{11}\text{C}$ ]-(+)-PHNO PET normative modelling. (A) Single-subject, whole-brain distributions of Z-scores estimated in the data before harmonisation (original data, on the left) and after spatial smoothing (in the centre) and Combat (on the right) harmonisation methods. This dataset combines scans of healthy controls acquired with either PET/CT (N=54) or PET/MR (N=23) scanners. (B) Welch two-sample t-tests comparing the first statistical moments (mean and standard deviation of the single-subject Z-score distributions between scanner types for each harmonisation modality (no harmonisation, harmonisation with spatial smoothing and harmonisation with Combat). Orange violin plots refer to PET/CT scans while blue violin plots refer

to PET/MR scans. Asterisks indicates significance. \* indicates  $p < 0.05$ , \*\* indicates  $p < 0.01$  and \*\*\*\* indicates  $p < 0.001$ .

When analysing between-scanner differences at the regional level (Supplementary figure 1), all harmonisation methods showed significant differences in the number of extreme deviations between scanner types ( $p < 0.01$ ), except Combat for the negative extreme deviations. Overall, Combat returned the smallest deltas, hence better harmonisation performances, for both positive extreme deviations (No harmonisation:  $\Delta_{MR-CT} = 4.30 \pm 2.21\%$ ; Smoothing:  $\Delta_{MR-CT} = -1.03 \pm 1.39\%$ ; Combat:  $\Delta_{MR-CT} = 0.44 \pm 1.04$ ) and negative extreme deviations (Original:  $\Delta_{MR-CT} = 6.02 \pm 2.30\%$ ; Smoothing:  $\Delta_{MR-CT} = 2.06 \pm 2.10\%$ ; Combat:  $\Delta_{MR-CT} = 0.06 \pm 0.97\%$ ) (Supplementary Figure 1).

## Validation of the optimal harmonisation strategy on [ $^{18}\text{F}$ ]FDOPA data

After harmonising [ $^{18}\text{F}$ ]FDOPA data with Combat, we estimated the voxel specific NMs, which explained up to 15% of the data variability although they do not converge to a solution for approximately 56% of the grey matter voxels (i.e.,  $\text{EXPV} < 0$ ; Supplementary Figure 2). All subsequent results will refer to NM data thresholded with the 3% variance mask, although the same analyses were repeated for the other masks (Supplementary Tables 1-12).

No clear scanner effects were apparent from the individual distributions of Z-scores in HC (Supplementary Figure 3). Similar to [ $^{11}\text{C}$ ]-(+)-PHNO PET NM, for FDOPA no significant scanner-related effect was found in terms of mean z-score distribution ( $F = 0.12$ ,  $p = \text{n.s.}$ ), while a significant residual effect was found for the standard deviation ( $F = 3.38$ ,  $p < 0.05$ ). Indicating that the harmonisation strategy was effectively applied to FDOPA.

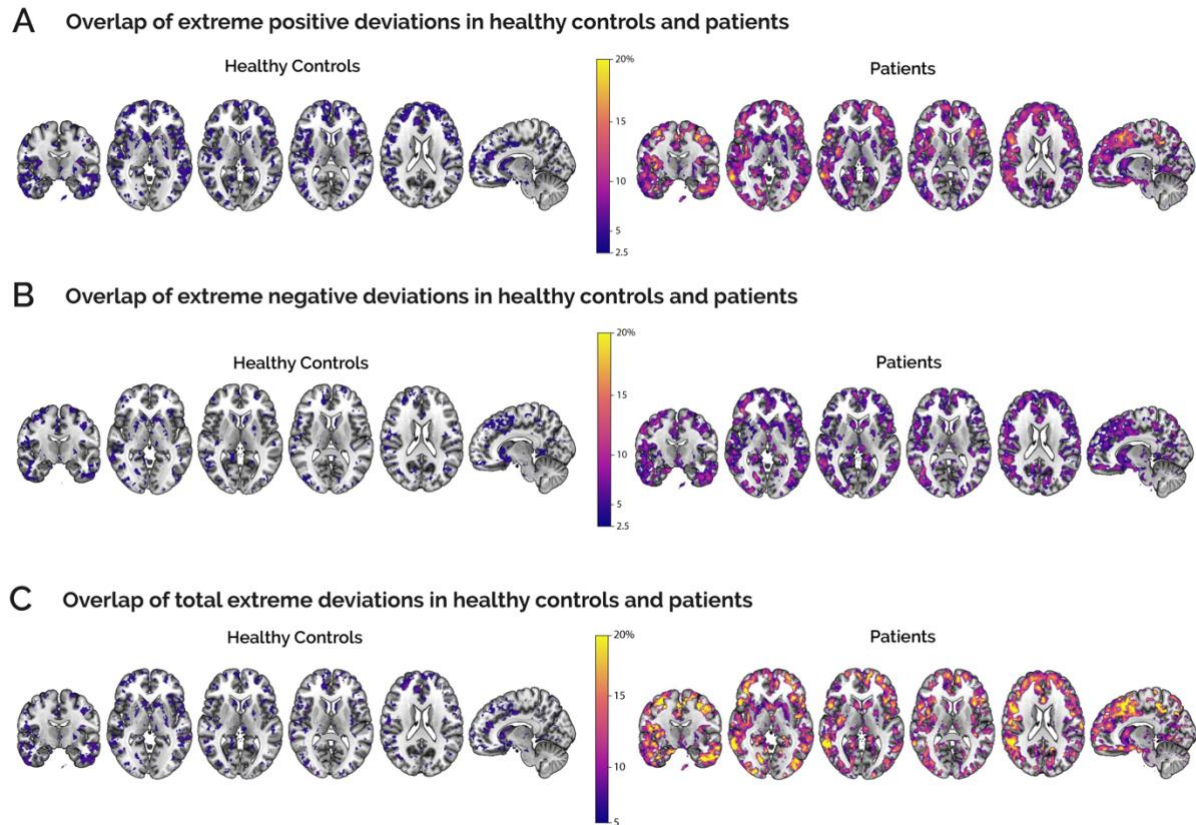
## NM application to [ $^{18}\text{F}$ ]FDOPA PET in psychosis

Grouped distributions of HC and patients showed small differences in the mean ( $|\Delta_{\mu}^{\text{HC-PAT}}| = 0.14$ ), and a greater difference in terms of the standard deviation ( $|\Delta_{\sigma}^{\text{HC-PAT}}| = 0.28$ ),

consistently higher for patients in both measures. The number of extremely deviating voxels (in percentage to the total voxels in the analysis mask), above the level of chance, shows increased RR for patients with all extreme deviation scores ( $RR_{\text{pos}}=2.19$ ,  $RR_{\text{neg}}=1.92$  and  $RR_{\text{tot}}=2.60$  for positive-, negative-, and total-extreme deviations respectively) (Supplementary Figure 2).

We hypothesised that the HC group would exhibit a relatively low percentage of individuals with extreme deviations, while conversely, patients would demonstrate a higher overlap in the percentage of individuals with extreme deviations from normality in specific brain regions. While HC reported very low percentages of individuals with overlapping extreme deviations (Figure 2), which aligns with the NM hypothesis for the reference cohort, patients exhibited distinct spatial patterns of extreme deviations, primarily in the cortex. These deviations were prominently present in the precentral and frontal gyri. However, it is important to emphasise that this pattern of extreme deviations also showed substantial sample heterogeneity, as the highest percentage of patients with comparable extreme deviations did not exceed 20%.



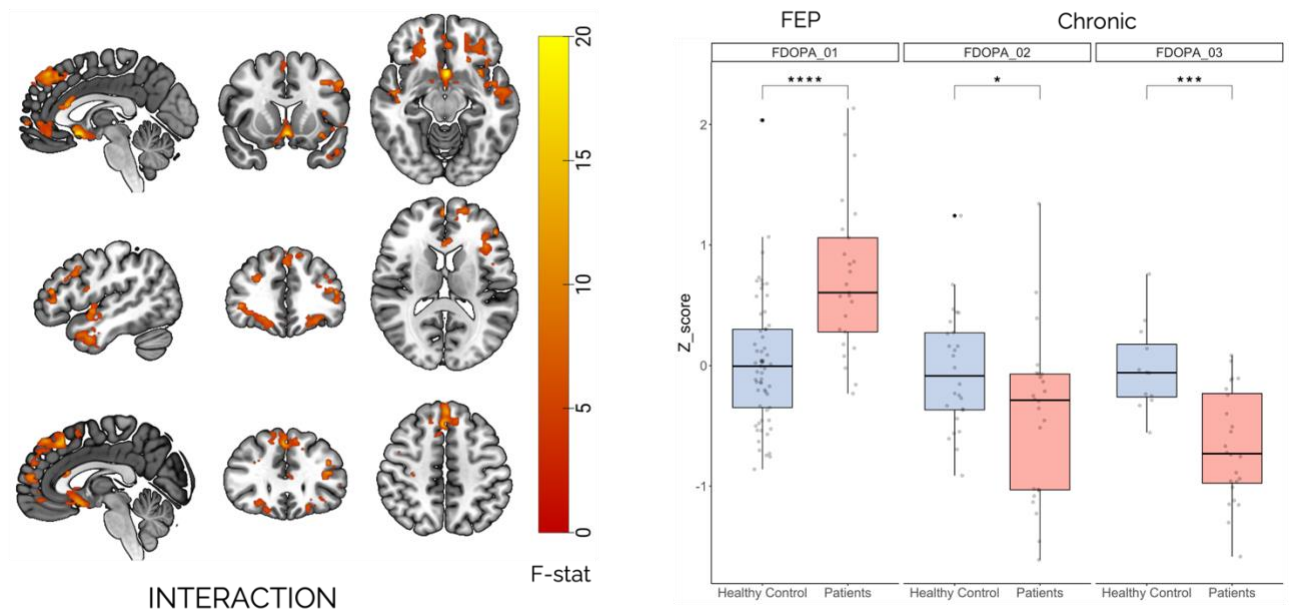


*Figure 2. Overlap maps of extreme deviations in HC and patients. Each map reports for each voxel the percentage of subjects with co-localised positive (top row), negative (middle row) and total (bottom row) extreme deviations. Controls (left) and psychosis patients (right) are reported separately. The figure is set at a threshold above the level of chance (2.5% for positive and negative deviations and 5% for total deviations). All overlap maps are superimposed on a standard MNI152 structural template.*

The voxel-wise two-way ANOVA performed on the non-thresholded Z-score maps revealed significant interactions between group and dataset ( $F_{\text{peak}}=21.71$ ,  $p_{\text{peak}}<0.001$  FWE corrected). This interaction spanned the temporal lobe, the medial- and superior-orbitofrontal, and the frontal lobe. The post hoc tests to evaluate between-group differences within each dataset highlighted greater positive deviations in the FEP as compared to the HC (FDOPA\_01,  $t = 4.92$ ,  $p < < 0.001$ ) and lower negative deviations in chronic patients as compared to HC in both FDOPA\_02 ( $t = 2.22$ ,  $p < 0.05$ ) and FDOPA\_03 ( $t = 4.62$ ,  $p < 0.001$ ).

We also found a group effect in the two-way ANOVA performed on the total extreme deviation maps ( $F_{\text{peak}}=34.36$ ,  $p<0.001$  FWE corrected) (Figure 3), exhibiting an increase in patients. This effect was localised in the left frontal orbital cortex, frontal pole, and inferior temporal cortex. Of note, only clusters including more than 50 voxels are reported.

### A) Voxelwise two-way ANOVA group-dataset interaction effect of average Z-score



### B) Voxelwise two-way ANOVA group effect of total extreme deviations

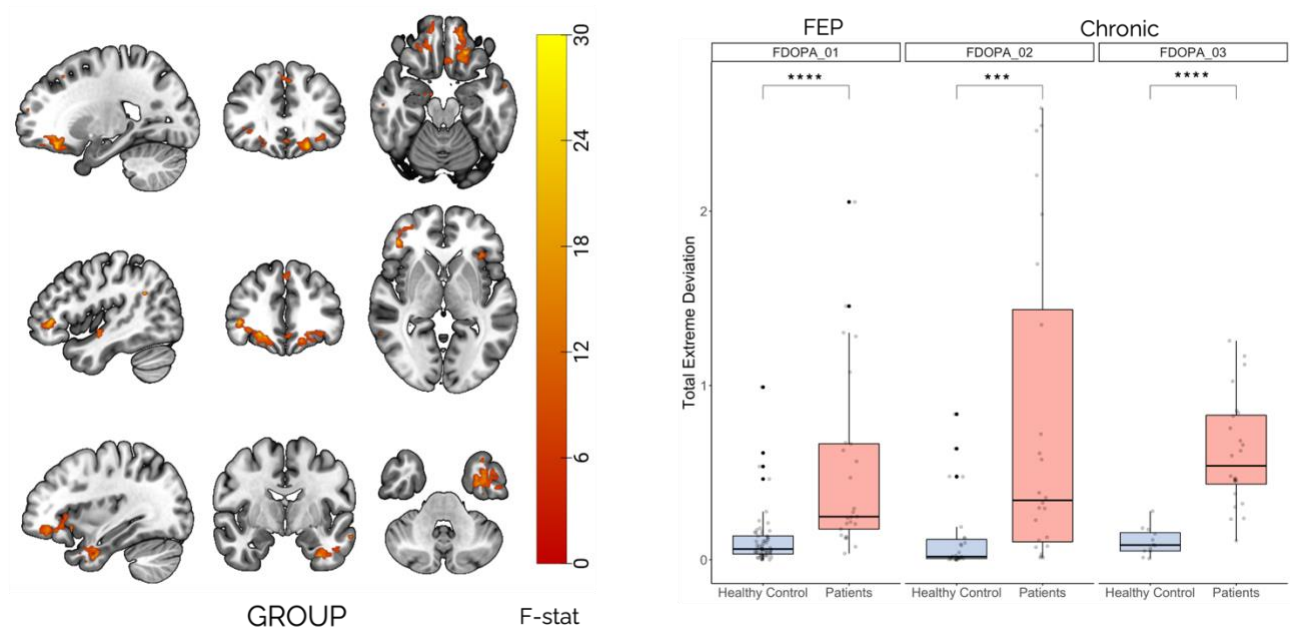
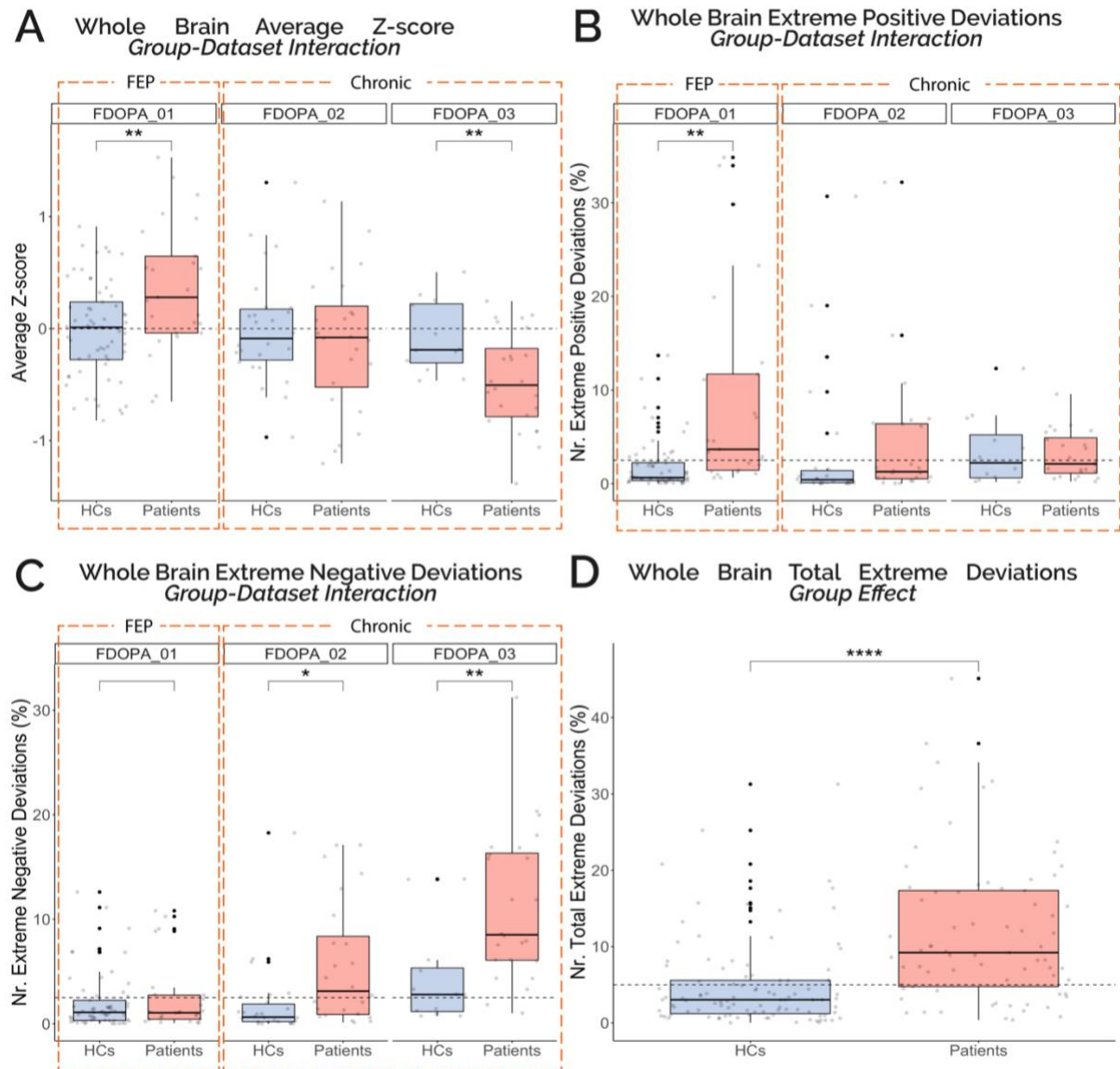


Figure 3. Voxel-wise ANOVA analysis of deviation scores, expressed as mean Z-score and total extreme deviations, for the [ $^{18}\text{F}$ ]FDOPA Normative Model. (A) Mean Z-score significant clusters (left)

*and post-hoc analyses, grouped by dataset, (right) between healthy controls and patients. (B) Total extreme deviations significant clusters (left) and post-hoc analyses, grouped by dataset, (right) between healthy controls and patients. \* indicates  $p < 0.05$ , \*\*\* indicates  $p < 0.001$  and \*\*\*\* indicates  $p < 0.001$ .*

The two-way ANOVAs conducted on the summary measures of the mean z-score and extreme positive, negative, and total deviations revealed a significant interaction between group and dataset for the mean Z-score ( $F = 8.43$ ,  $p < 0.001$ ; Figure 4A), extreme positive deviations ( $F = 5.90$ ,  $p < 0.01$ ; Figure 4B) and extreme negative deviations ( $F = 5.89$ ,  $p < 0.01$ , Figure 4C). Post hoc analyses showed an increase, in average Z-score in FEP ( $t = 3.08$ ,  $p < 0.01$ ), and a decrease in chronic patients ( $t = -3.05$ ,  $p < 0.01$ ) compared to HCs (Supplementary Table 5). Similarly, FEP was associated with an increase in extreme positive deviations compared to HCs ( $t = 3.23$ ,  $p < 0.01$ , Supplementary Table 6), whereas there was an increase of extreme negative deviations in chronic patients compared to HCs (FDOPA\_02,  $t = 2.67$ ,  $p < 0.05$ ; FDOPA\_03,  $t = 3.36$ ,  $p < 0.01$ ; Supplementary Table 7). The total extreme deviations, which combine positive and negative extreme values, did not exhibit a significant interaction between the group and dataset, although it did indicate a significant effect of the group ( $F = 37.52$ ,  $p < 0.001$ , Figure 4D, Supplementary Table 8), showing an increase in patients compared to HC.





*Figure 4. ANOVA analysis of individual deviation scores, expressed as mean Z-score and positive, negative, and total extreme deviations, for the [<sup>18</sup>F]FDOPA Normative Model. (A) Mean Z-score differences between healthy controls and patients grouped by dataset. (B) Difference of extreme positive deviations between healthy controls and patients grouped by dataset (C) Difference of extreme negative deviations between healthy controls and patients grouped by dataset. (D) Difference of total extreme deviations between group. Patients are composed of a mixture of first episode (FDOPA\_01) and chronic psychosis (FDOPA\_02, FDOPA\_03), depending on the dataset. Asterisks indicates significance, \* indicates  $p < 0.05$ , \*\* indicates  $p < 0.01$  and \*\*\*\* indicates  $p < 0.001$ .*

The imaging transcriptomic analysis conducted on the F-stat maps resulting from the significant group effect on the total extreme deviations, tested with two-way ANOVA

testing and representing abnormalities in patients regardless of the duration of psychosis, revealed several notable enriched gene sets. These included “structural constituent of ribosome” (NES=2.49,  $p_{FDR}<0.05$ ), “neurotransmitter receptor activity” (NES=2.18,  $p_{FDR}<0.05$ ) and “glutamate receptor activity” (NES=2.04,  $p_{FDR}<0.05$ ), as determined by the GO molecular function<sup>67,68</sup>.

## **Extending normative model validity to an independent patient cohort**

When visually compared with the HC of the independent datasets in terms of the percentage of individuals with extreme deviations (positive, negative, and total) within the group, the pattern of extreme deviations of the FDOPA\_04 dataset is consistent with that of the datasets of patients in FDOPA\_01, FDOPA\_02 and FDOPA\_03 (Supplementary Figures 5-7).

In terms of differences in the summary measures (mean z-score, and positive, negative, and total extreme deviations) between FDOPA\_04 patients and HC, we found significant increases in patients in terms of mean Z-score ( $t=-3.65$ ,  $p<0.001$ ), extreme positive deviation ( $t=-5.52$ ,  $p<0.001$ ) and total extreme deviations ( $t=-5.93$ ,  $p<0.001$ ). No significant differences were found in terms of negative extreme deviations ( $t=-1.14$ ,  $p=n.s.$ ).

## **Correlation between imaging-based deviations from normality and clinical symptoms**

We tested the hypothesis that there would be an association between the significant clusters from the voxel-wise two-way ANOVA performed on the extreme deviation maps and PANSS scores in both the FDOPA\_01 (FEP) and FDOPA\_04 (chronic schizophrenia) datasets. There were no significant correlations in the FDOPA\_01 dataset. In the FDOPA\_04 dataset, the mean Z-scores and total extreme deviations were significantly correlated with clinical scores (Table 2). Three individual clusters in FDOPA\_04 (with

$n_{\text{voxels}} \geq 50$ ) survived correction for multiple comparisons, showing significant correlations with total PANSS scores (Table 3).

	PANSS Positive		PANSS Negative		PANSS General		PANSS Total	
	FDOPA_01	FDOPA_04	FDOPA_01	FDOPA_04	FDOPA_01	FDOPA_04	FDOPA_01	FDOPA_04
<b>Average Z-score</b>	0.38	<b>0.36*</b>	0.38	<b>0.38*</b>	-0.06	<b>0.41*</b>	0.36	<b>0.43**</b>
<b>Extreme total deviations</b>	0.32	<b>0.36*</b>	0.31	<b>0.34*</b>	-0.09	<b>0.42*</b>	0.25	<b>0.44**</b>

*Table 2. Correlations between significant voxel-wise clusters and PANSS scores. The table shows Spearman correlation's rho and relative p-value for significant clusters resulting from voxel-wise two-way ANOVA (average Z-score, and  $|Z| > 2$ ). Correlations are separated for FDOPA\_01 (FEP patients, N=25) and FDOPA\_04 (chronic and FEP patients, N=36). \* indicates  $p < 0.05$ , \*\* indicates  $p < 0.01$*

Spearman correlation with PANSS Total scores			
	Correlation	Cluster size	Cluster Location
<b>Cluster 1</b>	<b>0.41*</b>	505	Frontal Orbital Cortex, Frontal Pole
<b>Cluster 2</b>	<b>0.36*</b>	421	Frontal Orbital Cortex, Frontal Pole, Inferior Frontal Gyrus
<b>Cluster 3</b>	<b>0.43*</b>	50	Middle Temporal Gyrus

*Table 3. Correlation analysis of PANSS total scores and individual voxel-wise clusters. Table shows Spearman correlation's rho and relative p-value for significant clusters between total extreme deviations and PANSS Total Scores in FDOPA\_04 (chronic patients), resulting from voxel-wise two-way ANOVA ( $|Z| > 2$ ). Cluster locations were obtained using FSL atlasquery from the Harvard-Oxford Cortical Atlas. \* indicates  $p < 0.05$ .*

Correlation of the summary deviations and PANSS scores revealed a significant correlation between total extreme deviations and PANSS negative in FEP patients (i.e., FDOPA\_01). The analysis of FDOPA\_04 reveals significant correlations between extreme positive deviations and PANSS general scores and between total extreme deviations and all PANSS scores (Table 3).

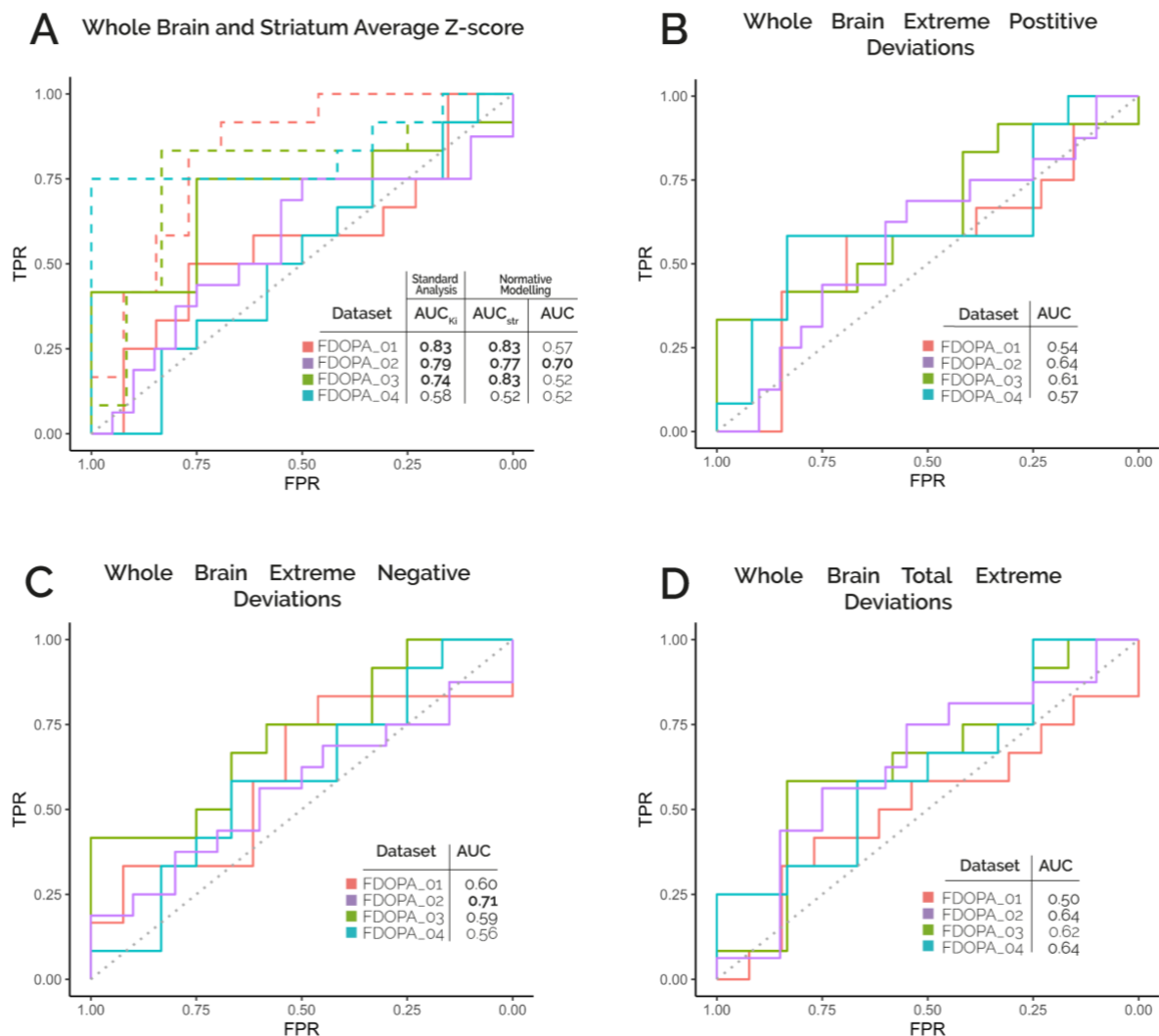
	PANSS Positive		PANSS Negative		PANSS General		PANSS Total	
	FDOPA_01	FDOPA_04	FDOPA_01	FDOPA_04	FDOPA_01	FDOPA_04	FDOPA_01	FDOPA_04
Average Z-score	0.28	0.25	0.30	0.16	-0.11	0.24	0.17	0.22
Extreme positive deviation	0.33	0.29	0.36	0.17	-0.12	<b>0.33*</b>	0.24	0.31
Extreme negative deviations	-0.07	-0.12	0.14	0.06	0.19	-0.01	0.11	0.03
Extreme total deviations	0.36	<b>0.33*</b>	<b>0.47*</b>	<b>0.39**</b>	-0.01	<b>0.44**</b>	0.30	<b>0.46**</b>

*Table 4. Correlation analysis of PANSS scores and summary measures. Table shows Spearman correlation's rho and relative p-value for whole brain summary measures (average Z-score,  $Z > 2$ ,  $Z < -2$ ,  $|Z| > 2$ ). Correlations are separated for FDOPA\_01 (FEP patients,  $N=25$ ) and FDOPA\_04 (chronic and FEP patients,  $N=36$ ). \* indicates  $p < 0.05$ , \*\* indicates  $p < 0.01$*

## Treatment Response in Patients

The analysis on the summary mean Z-score to evaluate patients' response to standard neuroleptics replicated the performance of the reference analysis (i.e.,  $K_i^{cer}$ ) and showed no significant difference between the two methods (as measured by the DeLong test) when focusing on the striatum only (Figure 5A). Here we obtained acceptable performances for FDOPA\_02 ( $AUC_{str}=0.77$ ,  $AUC_{ki}=0.79$ )<sup>70</sup>, and excellent performances for FDOPA\_01 ( $AUC_{str}=0.83$ ,  $AUC_{ki}=0.83$ ) and FDOPA\_03 ( $AUC_{str}=0.83$ ,  $AUC_{ki}=0.74$ )<sup>70</sup>.

However, when extending the analysis to the summary measures estimated at the whole brain, we did not reach acceptable classifications except for FDOPA\_02 (AUC=0.70, Figure 5A). Classification of treatment response measured with whole-brain deviation scores pointed at acceptable<sup>70</sup> performances only for FDOPA\_02 in terms of negative extreme deviations (AUC=0.72, Figure 5C).



*Figure 5. Clinical [<sup>18</sup>F]FDOPA Normative Model Treatment Response. ROC curves for the classification of treatment response (Responders vs Non-Responders/Resistant) using whole brain (solid lines) or striatal (dashed lines) individual average Z-score (A), whole brain positive extreme*

*deviations (B), negative extreme deviations (C) and total extreme deviations (D) for the different clinical datasets. Dashed diagonal indicates the level of chance in all plots. All scores refer to [<sup>18</sup>F]FDOPA patient data after Combat harmonization.*

## Discussion

This study applied normative modelling (NM) in PET imaging of the dopamine system in psychosis using several existing datasets. Two dopaminergic radiotracers ( $[^{11}\text{C}]$ -(+)-PHNO for  $D_{2/3}$  receptor and  $[^{18}\text{F}]$ FDOPA for dopamine synthesis) were used to construct normative models from reference healthy controls and applied to patients with psychosis. We demonstrate that NM is capable of explaining dopamine function variability arising from demographic factors and show that psychosis is associated with a differential pattern of extreme deviations.  $[^{18}\text{F}]$ FDOPA PET showed a high degree of patient heterogeneity, particularly in extra-striatal brain areas. In the striatum, deviation scores were able to classify treatment response, although with similar performance to standard kinetic modelling analysis. Overall, NM provides novel quantitative insights into the brain dopamine system in healthy individuals and people with psychosis, taking advantage of existing normative data built from multiple neuroimaging studies.

### ***Considerations on NM applied to molecular neuroimaging***

In our study, we first confirmed the necessity of data harmonisation to compile multi-centre, multi-scanner PET datasets with the required sample size for constructing a normative model of healthy cohorts. Our results indicated that Combat was the best harmonisation method to remove scanner and site effects. This finding aligns with the neuroimaging literature, where model-based harmonization methods are the preferred choices in PET<sup>71</sup> and other modalities<sup>32,33</sup>. It's worth noting that Combat assumes linearity in the scanner's effect, which may not hold true for tomographs with different geometries, axial fields of view, and gamma detectors. Recently, non-linear versions of Combat methods (e.g., Combat-GAM<sup>72</sup>) have been introduced, but they would require further testing with PET tracers exhibiting different kinetic properties. Regardless of the type of harmonization statistics used, our proposed approach relies on the existence of a group of healthy controls (HC) for scanner calibration. This results in a non-generalizable model without HC data. One solution, proposed in other neuroimaging modalities<sup>27,28,28,73</sup>, could be to incorporate the scanner directly into the normative model as a random effect using hierarchical modelling. While appealing for modelling flexibility, this methodology is

burdened by extreme computational times (reported to be up to ~24 hours per model<sup>27</sup>), making it unfeasible for our application (i.e., ~130K voxel-wise models).

Our normative model was able to explain a substantial amount of dopamine function variability (~30% in  $D_{2/3}$  receptor density, as measured by PHNO, and ~15% in dopamine synthesis capacity, as measured by FDOPA) using simple demographic factors from the population. This level of variance is lower than that reported in other modalities (40% for sMRI<sup>35</sup>, 29% for diffusion MRI<sup>74</sup>), which may relate to measurement of brain functional processes rather than structure. To allow generalizability, the covariates we selected were limited to general demographic factors (age and sex for [<sup>18</sup>F]FDOPA; age, sex, and BMI for [<sup>11</sup>C]-(+)-PHNO) were not specific to psychosis. Although the addition of more diagnosis-specific covariates and/or additional experimental factors (e.g., genetic traits), may have improved the normative model's performance, the use of ancillary measures (e.g., genetic tests or clinical assessments) might limit its potential clinical utility due to the necessity of having a PET scan and additional testing. Our model is therefore disease-agnostic.

### ***Considerations on NM applied to FDOPA PET imaging and patients with psychosis***

The application of an FDOPA NM to data acquired patients with psychosis provided several novel pathophysiological insights. Firstly, we observed that patients exhibit a higher percentage of extreme deviations than HCs. This observation is further supported by the elevated risk ratio (RR) for both positive and negative extreme deviations in patients compared to controls. Surprisingly, summary Z-score measures (calculated as individual whole-brain summary scores) exhibit a differential pattern consistent with the chronicity of the disease (i.e., first-episode psychosis – FEP, or chronic patients).

We observed that while there is a substantial overall difference due to the disease (as indicated by the ANOVA analysis of total extreme deviations), the pattern of positive and negative extreme deviations varies in different stages of the disease. For example, FEP patients show a higher number of positive extreme deviations than healthy controls and



chronic patients, while the latter show a higher number of negative extreme deviations than controls and FEP patients. This pattern is also evident in the average distribution scores and is confirmed by significant voxel-wise analyses (average Z-score and total extreme deviations).

Plausible explanations for this differential pattern are the effect of medications acting on the dopaminergic system, particularly  $D_{2/3}$  receptors, or grey matter reductions in chronic patients when compared to FEP<sup>75</sup>. Long-term antipsychotic treatment may up-regulate  $D_{2/3}$  receptors, which could potentially impact presynaptic dopamine synthesis capacity also through pre-synaptic autoregulatory mechanisms.

With the data available to our study, we cannot establish whether this differential pattern between FEP and chronic patients is attributed to variances in disease duration or is linked to medication effects. It is worth noting that each patient's experience with the disease is unique, and ideally, they should not be grouped with other patients for analytical purposes but rather considered as individual "clusters"<sup>76</sup>. Nevertheless, by employing NM, it would be feasible (provided the requisite data is available) to compare individual patients with consideration to their clinical and disease history.

Another interesting finding from our analyses is the substantial heterogeneity among patients, even though there is a consistent effect in all patients. This heterogeneity is particularly noticeable in analysis of the overlap of extreme deviations, where at most, 20% of patients co-localize in the same voxels. The co-localization of extreme deviations is more pronounced in patients compared to healthy controls (with a maximum expected overlap of 5%), indicating a common alteration of [<sup>18</sup>F]FDOPA PET signal in psychosis. These areas of extreme deviations are consistent with previous findings on NM applied to structural MRI, showing a maximum degree of overlap of 5-10%<sup>19,20</sup>. Future studies should explore of interlinks between structural and molecular changes at the individual level. Interestingly, [<sup>18</sup>F]FDOPA PET co-localizations of extreme deviations in psychosis occur in extra-striatal brain areas, contrary to the focus of most [<sup>18</sup>F]FDOPA studies on striatal areas<sup>38-47,49</sup>. Studies which have examined the cortical signal of [<sup>18</sup>F]FDOPA suggest elevations in dopamine synthesis capacity in cortical areas (e.g., posterior cingulate<sup>77</sup>, medial prefrontal cortex<sup>78</sup>) or correlations between dopamine synthesis capacity in cortical areas and clinical symptom severity (e.g., a positive correlation between PANSS

positive scores and  $K_i$  in the right temporal cortex<sup>79</sup>). These findings raise questions about alterations in aromatic L-amino acid decarboxylase (AADC) and whether patients have an unresponsive feedback mechanism of tyrosine hydroxylase, which may not down-regulate as in normal populations<sup>77</sup>. Early animal studies suggest that the [<sup>18</sup>F]FDOPA signal combines metabolic activity related to dopamine synthesis, storage, and metabolism<sup>80</sup>. Our imaging-transcriptomics analysis partially confirms these findings pointing towards an association between genetic pathways related to metabolism and [<sup>18</sup>F]FDOPA deviations measured in psychosis. All these findings suggest that cortical signal of [<sup>18</sup>F]FDOPA might be associated with increased cellular transport and/or protein metabolism, supporting the application of this biomarker in neuro-oncology<sup>81</sup>. Further studies are required to fully characterise the nature of the [<sup>18</sup>F]FDOPA signal in the human cortex, establishing the contributions of dopamine function (e.g., neuron density and changes in dopaminergic-mediated neuronal firing), metabolism and cellular transport to the overall signal. Lastly, there are genetic variants that are well-known to influence the dopamine system (e.g. 22q11.2 deletion/duplication<sup>40</sup> and COMT Val<sup>158</sup>Met polymorphism<sup>82</sup>) or that have been associated with the [<sup>18</sup>F]FDOPA PET signal (e.g. AS3MT/BORCS7 genetic variant<sup>83</sup>). These factors could drive some of the functional differences observed between patients and HCs and increase the inter-individual explained variance if included in the NM. Further analyses in a larger sample with full genetic information are warranted.

In our study, we explored the feasibility of investigating dopaminergic alterations of [<sup>18</sup>F]FDOPA without acquiring a matched set of healthy controls. With the analysis of the patient-only cohort (i.e., FDOPA\_04) and its comparison to the whole normative reference (i.e., all HCs) we obtained similar results as those in the other clinical datasets which had matched HCs in the normative reference, including the spatial pattern of the extreme deviations. This finding is particularly relevant, as routine clinical applications typically do not acquire matched controls for patients. The NM approach circumvents this issue by using a normative reference, accounting for covariates for comparison. Furthermore, the deviation scores of FDOPA\_04, especially total deviations, showed significant correlations with PANSS symptoms, both as whole-brain summary scores and significant cluster measures.

Interestingly, by analysing individual clusters within the cluster signal, we found that the clusters correlating with total PANSS symptoms were in the frontal and temporal gyri. However, no significant correlations were observed in the smaller FDOPA\_01 dataset, which may relate to lower statistical power. This framework, along with automatic standardised frameworks for kinetic modelling of [ $^{18}\text{F}$ ]FDOPA<sup>49</sup>, represents a step forward in making PET parametric imaging a quantitative companion tool for clinical management of psychosis patients.

Finally, we analysed the potential of using NM-derived scores for classifying treatment response in patients (i.e., those who will respond to treatment at follow-up based on baseline scores) and compared these findings to previously published results<sup>38</sup>. Interestingly, when focusing on striatum NM, we replicated the prediction performances of the reference standard<sup>38</sup>, while poor classification performances were obtained when extending the method to the whole brain. This outcome aligns with expectations, as the molecular target of action of both antipsychotic medications is striatal dopamine  $D_{2/3}$  receptors<sup>84,85</sup>. In fact, the striatal average Z-scores (as defined by the Hammersmith atlas and where the model converges) were highly correlated with  $K_i$  estimates in all patient datasets ( $0.93 < R < 0.97$ ,  $p < 0.001$ , Supplementary Figure 8), providing thus an alternative view of the same information. In line with this, the distributions of extreme positive and total extreme deviations in the striatal region revealed group differences, indicating an elevation in the dopamine signal (Supplementary Figure 9).

### **Limitations**

This work is subject to several limitations. Firstly, the modelling process is influenced by the harmonization method and may not account for non-linearities within the system, either in harmonization or modelling. Despite our efforts to harmonize data from different clinical sites, residual differences still exist in the datasets when assessed with statistics of the second order or above (e.g., skewness, kurtosis). Secondly, applying NM at the voxel level can result in underperformance of the model due to high noise and poor signal. Consequently, there are a considerable number of voxels where the model fails to converge. This issue could potentially be addressed by adjusting the MCMC settings or

by incorporating a more extensive set of covariates to describe the neuroimaging target variable. While NM mapping provides a topological description of patient deviation, region-based modelling remains a valid compromise between spatial resolution and method performance. Additionally, the deviation scores might be sensitive to motion<sup>86</sup>, however when assessing for our data there were no significant correlations in HCs nor in patients (Supplementary Figure 10). Moreover, the heterogeneity of the clinical sample used in the analysis and the presence of missing clinical information for some individuals in the cohort are likely to have affected statistical power. Lastly, our result could be influenced by the choice of EXPV threshold used for masking, although similar results were obtained for the other threshold tested (Supplementary Tables 1-12).

## **Conclusions**

In conclusion, the NM framework can be successfully applied to molecular neuroimaging (i.e., PET and SPECT) after proper harmonisation of scanner effects. Moreover, with the NM model, we can assess a differential pattern of deviations likely attributable to the chronicity of the disease and compare a patient-only cohort to the normative reference to gain mechanistic insights and advance toward a quantitative and biological understanding of psychosis. Additionally, we are able to replicate the findings of traditional cross-sectional studies and performances with standard analytical approaches. While the focus of this work was on the presynaptic dopaminergic system (as measured by [<sup>18</sup>F]FDOPA) and on the post-synaptic dopaminergic system (as measured by [<sup>11</sup>C]-(+)-PHNO) in the context of psychosis, we believe that after selecting appropriate covariates, this methodology can be applied to any molecular target measured by PET or SPECT neuroimaging.

## **Acknowledgements**

This study represents independent research funded by the NIHR Maudsley Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London. The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care. A.G. is supported by the

KCL-funded CDT in Data-Driven Health; this represents independent research partly funded by the NIHR Maudsley's Biomedical Research Centre (BRC) at the South London and Maudsley NHS Foundation Trust and partly funded by GSK. D.M., G.N., R.E., O.D., F.T., and S.C.R.W. are supported by the NIHR Maudsley's Biomedical Research Centre at the South London and Maudsley NHS Trust. MV is supported by the Italian National Centre for HPC, BIG DATA AND QUANTUM COMPUTING (Project no. CN00000013 CN1), the PNR Italian National Grant DIGITAL LIFELONG PREVENTION (Project no PNC0000002\_DARE), and by Wellcome Trust Digital Award (no. 215747/Z/19/Z). M.d.G. is an employee of GSK, GSK had no role in the design of this study.

## Conflict of Interests

M.d.G. is an employee of GSK, GSK had no role in the design of this study. R.A.M. has received speaker/consultancy fees from Karuna, Janssen, Boehringer Ingelheim, and Otsuka, and co-directs a company that designs digital resources to support treatment of mental illness. F.B. has received consulting fees from Petalouda Therapeutics and has been an employee at Compass Pathways. AE has received consulting fees from Leal Therapeutics.

## Supplementary Information

Supplementary information is available at MP's website.

## References

1. Ellison-Wright, I. & Bullmore, E. Meta-analysis of diffusion tensor imaging studies in schizophrenia. *Schizophr. Res.* **108**, 3–10 (2009).

2. Dong, D., Wang, Y., Chang, X., Luo, C. & Yao, D. Dysfunction of Large-Scale Brain Networks in Schizophrenia: A Meta-analysis of Resting-State Functional Connectivity. *Schizophr. Bull.* **44**, 168–181 (2018).
3. Townsend, L. *et al.* Brain glucose metabolism in schizophrenia: a systematic review and meta-analysis of 18FDG-PET studies in schizophrenia. *Psychol. Med.* 1–18 (2022)  
doi:10.1017/S003329172200174X.
4. Marques, T. R. *et al.* Neuroinflammation in schizophrenia: meta-analysis of in vivo microglial imaging studies. *Psychol. Med.* **49**, 2186–2196 (2019).
5. Howes, O. D. *et al.* The Nature of Dopamine Dysfunction in Schizophrenia and What This Means for Treatment: Meta-analysis of Imaging Studies. *Arch. Gen. Psychiatry* **69**, 776–786 (2012).
6. Howes, O. *et al.* Mechanisms underlying psychosis and antipsychotic treatment response in schizophrenia: insights from PET and SPECT imaging. *Curr. Pharm. Des.* **15**, 2550–2559 (2009).
7. Howes, O. D. & Kapur, S. The Dopamine Hypothesis of Schizophrenia: Version III—The Final Common Pathway. *Schizophr. Bull.* **35**, 549–562 (2009).
8. Demjaha, A., Murray, R. M., McGuire, P. K., Kapur, S. & Howes, O. D. Dopamine synthesis capacity in patients with treatment-resistant schizophrenia. *Am. J. Psychiatry* **169**, 1203–1210 (2012).
9. Howes, O. D. *et al.* Dopamine synthesis capacity before onset of psychosis: a prospective [18F]-DOPA PET imaging study. *Am. J. Psychiatry* **168**, 1311–1317 (2011).
10. Howes, O. *et al.* Progressive increase in striatal dopamine synthesis capacity as patients develop psychosis: a PET study. *Mol. Psychiatry* **16**, 885–886 (2011).

11. Javitt, D. C. Glutamate and schizophrenia: phencyclidine, N-methyl-D-aspartate receptors, and dopamine-glutamate interactions. *Int. Rev. Neurobiol.* **78**, 69–108 (2007).
12. Javitt, D. C. Glutamatergic theories of schizophrenia. *Isr. J. Psychiatry Relat. Sci.* **47**, 4–16 (2010).
13. Bethlehem, R. a. I. *et al.* Brain charts for the human lifespan. *Nature* **604**, 525–533 (2022).
14. Marquand, A. F., Rezek, I., Buitelaar, J. & Beckmann, C. F. Understanding Heterogeneity in Clinical Cohorts Using Normative Models: Beyond Case-Control Studies. *Biol. Psychiatry* **80**, 552–561 (2016).
15. Kapur, S., Phillips, A. G. & Insel, T. R. Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it? *Mol. Psychiatry* **17**, 1174–1179 (2012).
16. Cole, T. J. The development of growth references and growth charts. *Ann. Hum. Biol.* **39**, 382–394 (2012).
17. Marquand, A. F. *et al.* Conceptualizing mental disorders as deviations from normative functioning. *Mol. Psychiatry* **24**, 1415–1424 (2019).
18. Rutherford, S. *et al.* Charting brain growth and aging at high spatial precision. *eLife* **11**, e72904 (2022).
19. Wolfers, T. *et al.* Mapping the Heterogeneous Phenotype of Schizophrenia and Bipolar Disorder Using Normative Models. *JAMA Psychiatry* **75**, 1146 (2018).
20. Wolfers, T. *et al.* Replicating extensive brain structural heterogeneity in individuals with schizophrenia and bipolar disorder. *Hum. Brain Mapp.* **42**, 2546–2555 (2021).
21. Bethlehem, R. A. I. *et al.* A normative modelling approach reveals age-atypical cortical thickness in a subgroup of males with autism spectrum disorder. *Commun. Biol.* **3**, 1–10 (2020).

22. Rutherford, S. *et al.* The normative modeling framework for computational psychiatry. *Nat. Protoc.* **17**, 1711–1734 (2022).
23. Antoniadou, M. *et al.* Personalized Estimates of Brain Structural Variability in Individuals With Early Psychosis. *Schizophr. Bull.* **47**, 1029–1038 (2021).
24. Wolfers, T. *et al.* Individual differences v. the average patient: mapping the heterogeneity in ADHD using normative models. *Psychol. Med.* **50**, 314–323 (2020).
25. Chen, C.-L. *et al.* Advanced brain aging in multiple system atrophy compared to Parkinson’s disease. *NeuroImage Clin.* **34**, 102997 (2022).
26. Elad, D. *et al.* Improving the predictive potential of diffusion MRI in schizophrenia using normative models—Towards subject-level classification. *Hum. Brain Mapp.* **42**, 4658–4670 (2021).
27. Bayer, J. M. M. *et al.* Accommodating site variation in neuroimaging data using normative and hierarchical Bayesian models. 2021.02.09.430363 Preprint at <https://doi.org/10.1101/2021.02.09.430363> (2021).
28. Kia, S. M. *et al.* Hierarchical Bayesian Regression for Multi-Site Normative Modeling of Neuroimaging Data. *ArXiv200512055 Cs Stat* (2020).
29. Norgaard, M. *et al.* *PET-BIDS, an extension to the brain imaging data structure for positron emission tomography.*  
<http://biorxiv.org/lookup/doi/10.1101/2021.06.16.448390> (2021)  
doi:10.1101/2021.06.16.448390.
30. Knudsen, G. M. *et al.* Guidelines for the content and format of PET brain data in publications and archives: A consensus paper. *J. Cereb. Blood Flow Metab.* **40**, 1576–1585 (2020).



31. Thompson, P. M. *et al.* The ENIGMA Consortium: large-scale collaborative analyses of neuroimaging and genetic data. *Brain Imaging Behav.* **8**, 153–182 (2014).
32. Fortin, J.-P. *et al.* Harmonization of cortical thickness measurements across scanners and sites. *NeuroImage* **167**, 104–120 (2018).
33. Fortin, J.-P. *et al.* Harmonization of multi-site diffusion tensor imaging data. *NeuroImage* **161**, 149–170 (2017).
34. Joshi, A., Koeppe, R. A. & Fessler, J. A. Reducing between scanner differences in multi-center PET studies. *NeuroImage* **46**, 154–159 (2009).
35. Bayer, J. M. M. *et al.* Site effects how-to & when: an overview of retrospective techniques to accommodate site effects in multi-site neuroimaging analyses. Preprint at <https://doi.org/10.31234/osf.io/mpufv> (2022).
36. Dahoun, T. *et al.* The relationship between childhood trauma, dopamine release and dexamphetamine-induced positive psychotic symptoms: a [11C]-(+)-PHNO PET study. *Transl. Psychiatry* **9**, (2019).
37. Selvaggi, P. *et al.* Increased cerebral blood flow after single dose of antipsychotics in healthy volunteers depends on dopamine D2 receptor density profiles. *NeuroImage* **188**, 774–784 (2019).
38. Veronese, M. *et al.* A potential biomarker for treatment stratification in psychosis: evaluation of an [18F] FDOPA PET imaging approach. *Neuropsychopharmacology* **46**, 1122–1132 (2021).
39. Jauhar, S. *et al.* A Test of the Transdiagnostic Dopamine Hypothesis of Psychosis Using Positron Emission Tomographic Imaging in Bipolar Affective Disorder and Schizophrenia. *JAMA Psychiatry* **74**, 1206–1213 (2017).

40. Rogdaki, M. *et al.* Striatal dopaminergic alterations in individuals with copy number variants at the 22q11.2 genetic locus and their implications for psychosis risk: a [18F]-DOPA PET study. *Mol. Psychiatry* 1–12 (2021) doi:10.1038/s41380-021-01108-y.
41. Bloomfield, M. A. P. *et al.* Dopaminergic Function in Cannabis Users and Its Relationship to Cannabis-Induced Psychotic Symptoms. *Biol. Psychiatry* **75**, 470–478 (2014).
42. Jauhar, S. *et al.* Determinants of treatment response in first-episode psychosis: an 18F-DOPA PET study. *Mol. Psychiatry* **24**, 1502–1512 (2019).
43. Kim, E. *et al.* Presynaptic Dopamine Capacity in Patients with Treatment-Resistant Schizophrenia Taking Clozapine: An [18F]DOPA PET Study. *Neuropsychopharmacol. Off. Publ. Am. Coll. Neuropsychopharmacol.* **42**, 941–950 (2017).
44. Egerton, A. *et al.* Dopamine and Glutamate in Antipsychotic-Responsive Compared With Antipsychotic-Nonresponsive Psychosis: A Multicenter Positron Emission Tomography and Magnetic Resonance Spectroscopy Study (STRATA). *Schizophr. Bull.* **47**, 505–516 (2021).
45. Egerton, A. *et al.* Presynaptic Striatal Dopamine Dysfunction in People at Ultra-high Risk for Psychosis: Findings in a Second Cohort. *Biol. Psychiatry* **74**, 106–112 (2013).
46. Jauhar, S. *et al.* The Effects of Antipsychotic Treatment on Presynaptic Dopamine Synthesis Capacity in First-Episode Psychosis: A Positron Emission Tomography Study. *Biol. Psychiatry* **85**, 79–87 (2019).
47. Howes, O. D. *et al.* Dopaminergic Function in the Psychosis Spectrum: An [18F]-DOPA Imaging Study in Healthy Individuals With Auditory Hallucinations. *Schizophr. Bull.* **39**, 807–814 (2013).
48. Innis, R. B. *et al.* Consensus Nomenclature for in vivo Imaging of Reversibly Binding Radioligands. *J. Cereb. Blood Flow Metab.* **27**, 1533–1539 (2007).

49. Nordio, G. *et al.* An automatic analysis framework for FDOPA PET neuroimaging. *J. Cereb. Blood Flow Metab.* 0271678X231168687 (2023)  
doi:10.1177/0271678X231168687.
50. Kay, S. R., Fiszbein, A. & Opler, L. A. The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophr. Bull.* **13**, 261–276 (1987).
51. Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W. & Smith, S. M. FSL. *NeuroImage* **62**, 782–790 (2012).
52. Smith, S. M. *et al.* Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage* **23**, S208–S219 (2004).
53. Woolrich, M. W. *et al.* Bayesian analysis of neuroimaging data in FSL. *NeuroImage* **45**, S173–S186 (2009).
54. Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118–127 (2007).
55. NRM 2021 Abstract Booklet. *J. Cereb. Blood Flow Metab.* **41**, 11–309 (2021).
56. Matuskey, D. *et al.* Age-related changes in binding of the D2/3 receptor radioligand [11C](+)PHNO in healthy volunteers. *NeuroImage* **130**, 241–247 (2016).
57. Gaiser, E. C. *et al.* Elevated Dopamine D2/3 Receptor Availability in Obese Individuals: A PET Imaging Study with [11C](+)PHNO. *Neuropsychopharmacology* **41**, 3042–3050 (2016).
58. Kassambara, A. rstatix: Pipe-Friendly Framework for Basic Statistical Tests. (2021).
59. Hammers, A. *et al.* Three-dimensional maximum probability atlas of the human brain, with particular reference to the temporal lobe. *Hum. Brain Mapp.* **19**, 224–247 (2003).
60. Altman, D. G. *Practical Statistics for Medical Research*. (Chapman and Hall/CRC, 1990).  
doi:10.1201/9780429258589.

61. Winkler, A. M., Ridgway, G. R., Webster, M. A., Smith, S. M. & Nichols, T. E. Permutation inference for the general linear model. *NeuroImage* **92**, 381–397 (2014).
62. Smith, S. M. & Nichols, T. E. Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. *NeuroImage* **44**, 83–98 (2009).
63. Salimi-Khorshidi, G., Smith, S. M. & Nichols, T. E. Adjusting the effect of nonstationarity in cluster-based and TFCE inference. *NeuroImage* **54**, 2006–2019 (2011).
64. Martins, D. *et al.* Imaging transcriptomics: Convergent cellular, transcriptomic, and molecular neuroimaging signatures in the healthy adult human brain. *Cell Rep.* **37**, 110173 (2021).
65. Giacomel, A. *et al.* The Imaging Transcriptomics Toolbox. (2022)  
doi:10.5281/zenodo.6364963.
66. Giacomel, A. *et al.* Integrating neuroimaging and gene expression data using the imaging transcriptomics toolbox. *STAR Protoc.* **3**, 101315 (2022).
67. Ashburner, M. *et al.* Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).
68. The Gene Ontology Consortium *et al.* The Gene Ontology knowledgebase in 2023. *Genetics* **224**, iyad031 (2023).
69. Robin, X. *et al.* pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* **12**, 77 (2011).
70. Mandrekar, J. N. Receiver Operating Characteristic Curve in Diagnostic Test Assessment. *J. Thorac. Oncol.* **5**, 1315–1316 (2010).
71. Orhac, F. *et al.* A Guide to ComBat Harmonization of Imaging Biomarkers in Multicenter Studies. *J. Nucl. Med.* **63**, 172–179 (2022).

72. Pomponio, R. *et al.* Harmonization of large MRI datasets for the analysis of brain imaging patterns throughout the lifespan. *NeuroImage* **208**, 116450 (2020).
73. Kia, S. M. *et al.* Federated Multi-Site Normative Modeling using Hierarchical Bayesian Regression. 2021.05.28.446120 Preprint at <https://doi.org/10.1101/2021.05.28.446120> (2021).
74. Fraza, C. J., Dinga, R., Beckmann, C. F. & Marquand, A. F. Warped Bayesian Linear Regression for Normative Modelling of Big Data. *bioRxiv* 2021.04.05.438429 (2021) doi:10.1101/2021.04.05.438429.
75. Liu, N. *et al.* Characteristics of gray matter alterations in never-treated and treated chronic schizophrenia patients. *Transl. Psychiatry* **10**, 1–10 (2020).
76. Meyer-Lindenberg, A. The non-ergodic nature of mental health and psychiatric disorders: implications for biomarker and diagnostic research. *World Psychiatry* **22**, 272–274 (2023).
77. Elkashef, A. M. *et al.* 6-18F-DOPA PET study in patients with schizophrenia. *Psychiatry Res. Neuroimaging* **100**, 1–11 (2000).
78. Lindström, L. H. *et al.* Increased dopamine synthesis rate in medial prefrontal cortex and striatum in schizophrenia indicated by L-( $\beta$ -11C) DOPA and PET. *Biol. Psychiatry* **46**, 681–688 (1999).
79. Nozaki, S. *et al.* Regional dopamine synthesis in patients with schizophrenia using L-[ $\beta$ -11C]DOPA PET. *Schizophr. Res.* **108**, 78–84 (2009).
80. Garnett, E. S., Firnau, G., Chan, P. K. H., Sood, S. & Belbeck, L. W. [18F]Fluoro-dopa, an analogue of dopa, and its use in direct external measurements of storage, degradation, and turnover of intracerebral dopamine. *Proc. Natl. Acad. Sci.* **75**, 464–467 (1978).

81. Somme, F., Bender, L., Namer, I. J., Noël, G. & Bund, C. Usefulness of 18F-FDOPA PET for the management of primary brain tumors: a systematic review of the literature. *Cancer Imaging* **20**, 70 (2020).
82. Ira, E., Zanoni, M., Ruggeri, M., Dazzan, P. & Tosato, S. COMT, neuropsychological function and brain structure in schizophrenia: a systematic review and neurobiological interpretation. *J. Psychiatry Neurosci.* **38**, 366–380 (2013).
83. D’Ambrosio, E. *et al.* The effect of a genetic variant at the schizophrenia associated AS3MT/BORCS7 locus on striatal dopamine function: A PET imaging study. *Psychiatry Res. Neuroimaging* **291**, 34–41 (2019).
84. Rampino, A. *et al.* Antipsychotic Drug Responsiveness and Dopamine Receptor Signaling; Old Players and New Prospects. *Front. Psychiatry* **9**, (2019).
85. Aringhieri, S. *et al.* Molecular targets of atypical antipsychotics: From mechanism of action to clinical differences. *Pharmacol. Ther.* **192**, 20–41 (2018).
86. Cirstian, R. *et al.* Objective QC for diffusion MRI data: artefact detection using normative modelling. 2023.06.30.546837 Preprint at <https://doi.org/10.1101/2023.06.30.546837> (2023).

## **TABLES AND FIGURES**



Table 1. Study datasets.

<b>[<sup>11</sup>C]-(+)-PHNO Datasets</b>										
Group	Age mean (SD) [years]	Sex n (%)	BMI mean (SD) [kg/m <sup>2</sup> ]	N. controls	N. patients	N. resp/non-resp (resistant)	Patients	Scanner (scanner type)	Imaging Site	Clinical Scores
PET/CT	26.52 (7.87)	22 F (41%)/ 32 M (59%)	23.26 (2.78)	54	n.a.	n.a.	n.a.	Hi-Rez Biograph 6 (PET/CT)	Invicro	n.a.
PET/MR	24.39 (4.90)	12 F (52%)/ 11 M (48%)	22.47 (2.84)	23	n.a.	n.a.	n.a.	Biograph mMR (PET/MR)	Invicro	n.a.
<b>[<sup>18</sup>F]FDOPA Datasets</b>										
Group	Age mean (SD) [years]	Sex n (%)	BMI mean (SD) [kg/m <sup>2</sup> ]	N. controls	N. patients	N. resp/non-resp (or resistant)	Patients	Scanner (scanner type)	Imaging Site	Clinical Scores
FDOPA_01	26.25 (5.18)	32 F (36.7%)/ 55 M (63.21%)	n.a.	61	25	13/12	First Episode Patients (FEP)	Hi-Rez Biograph 6 (PET/CT)	Invicro	Yes
FDOPA_02	29.00 (8.70)	17 F (34.0%)/ 33 M (66.0%)	n.a.	26	24	12/12	Chronic	Biograph 40 TruePoint (PET/CT)	South Korea	No
FDOPA_03	44.52 (9.97)	21 F (58.88%)/	n.a.	12	24	12/12	Chronic	ECAT HR+ 962 (PET/CT)	MRC Cyclotron Unit	No

		13 M (36.22%)								
<b>FDOPA_04</b>	30.63 (10.91)	8 F (22.22%)/ 28 M (77.77%)	n.a.	n.a.	36	20/16	Mixed (FEPs + chronic)	Hi-Rez Biograph 6 (PET/CT)	Invicro	Yes
<b>FDOPA_05</b>	29.42 (6.09)	14 F (38.89%)/ 22 M (61.11%)	n.a.	36	n.a.	n.a.	n.a.	Biograph TruePoint 6 (PET/CT)	Invicro	n.a.
<b>FDOPA_06</b>	24.43 (4.65)	1 F (14.3%)/ 6 M (85.7%)	n.a.	7	n.a.	n.a.	n.a.	ECAT EXACT 3D (PET/CT)	MRC Cyclotron Unit	n.a.

\*n.a. indicates where the data are not available. First Episode Patients (FEP) indicates patients scanned shortly after the first psychosis episode, while chronic patients are patients who have undergone several rounds of standard line antipsychotics treatment but have not responded to treatment.

	PANSS Positive		PANSS Negative		PANSS General		PANSS Total	
	FDOPA_01	FDOPA_04	FDOPA_01	FDOPA_04	FDOPA_01	FDOPA_04	FDOPA_01	FDOPA_04
Average Z-score	0.38	0.36*	0.38	0.38*	-0.06	0.41*	0.36	0.43**
Extreme total deviations	0.32	0.36*	0.31	0.34*	-0.09	0.42*	0.25	0.44**

*Table 2. Correlation analysis of PANSS scores and significant voxel-wise clusters. Table shows Spearman correlation's rho and relative p-value for significant clusters resulting from voxel-wise two-way ANOVA (average Z-score, and  $|Z|>2$ ). Correlations are separated for FDOPA\_01 (FEP patients, N=25) and FDOPA\_04 (chronic and FEP patients, N=36). \* indicates  $p<0.05$ , \*\* indicates  $p<0.01$*

Spearman correlation with PANSS Total scores		
	Correlation	Cluster size
Cluster 1	0.41*	505
Cluster 2	0.36*	421
Cluster 3	0.43*	50

*Table 3. Correlation analysis of PANSS total scores and individual voxel-wise clusters. Table shows Spearman correlation's rho and relative p-value for significant clusters resulting from voxel-wise two-way ANOVA ( $|Z|>2$ ). \* indicates  $p<0.05$ .*

	PANSS Positive		PANSS Negative		PANSS General		PANSS Total	
	FDOPA_01	FDOPA_04	FDOPA_01	FDOPA_04	FDOPA_01	FDOPA_04	FDOPA_01	FDOPA_04
Average Z-score	0.28	0.25	0.30	0.16	-0.11	0.24	0.17	0.22
Extreme positive deviation	0.33	0.29	0.36	0.17	-0.12	0.33*	0.24	0.31
Extreme negative deviations	-0.07	-0.12	0.14	0.06	0.19	-0.01	0.11	0.03

<b>Extreme total deviations</b>	0.36	0.33*	0.47*	0.39**	-0.01	0.44**	0.30	0.46**
---------------------------------	------	-------	-------	--------	-------	--------	------	--------

*Table 4. Correlation analysis of PANSS scores and summary measures. Table shows Spearman correlation's rho and relative p-value for whole brain summary measures (average Z-score,  $Z > 2$ ,  $Z < -2$ ,  $|Z| > 2$ ). Correlations are separated for FDOPA\_01 (FEP patients, N=25) and FDOPA\_04 (chronic and FEP patients, N=36). \* indicates  $p < 0.05$ , \*\* indicates  $p < 0.01$*

*Figure 2. Harmonisation results of [<sup>11</sup>C]-(+)-PHNO PET normative modelling. (A) Single-subject, whole-brain distributions of Z-scores estimated in the data before harmonisation (original data, on the left) and after spatial smoothing (in the centre) and Combat (on the right) harmonisation methods. This dataset combines scans of healthy controls acquired with either PET/CT (N=54) or PET/MR (N=23) scanners. (B) Welch two-sample t-tests comparing the first statistical moments (mean and standard deviation of the single-subject Z-score distributions between scanner types for each harmonisation modality (no harmonisation, harmonisation with spatial smoothing and harmonisation with Combat). Orange violin plots refer to PET/CT scans while blue violin plots refer to PET/MR scans. Asterisks indicates significance. \* indicates  $p < 0.05$ , \*\* indicates  $p < 0.01$  and \*\*\*\* indicates  $p < 0.001$ .*

*Figure 2. Overlap maps of extreme deviations in HC and patients. Each map reports for each voxel the percentage of subjects with co-localise positive (top row), negative (middle row) and total (bottom row) extreme deviations. Controls (left) and psychosis patients (right) are reported separately. The figure is thresholded above the level of chance (2.5% for positive and negative deviations and 5% for total deviations). All overlap maps are superimposed on a standard MNI152 structural template.*

*Figure 3. Voxel-wise ANOVA analysis of deviation scores, expressed as mean Z-score and total extreme deviations, for the [<sup>18</sup>F]FDOPA Normative Model. (A) Mean Z-score significant clusters (left) and post-hoc analyses, grouped by dataset, (right) between healthy controls and patients. (B) Total extreme deviations significant clusters (left) and post-hoc analyses, grouped by dataset, (right) between healthy controls and patients. \* indicates  $p < 0.05$ , \*\*\* indicates  $p < 0.001$  and \*\*\*\* indicates  $p < 0.001$ .*

*Figure 4. ANOVA analysis of individual deviation scores, expressed as mean Z-score and positive, negative, and total extreme deviations, for the [<sup>18</sup>F]FDOPA Normative Model. (A) Mean Z-score differences between healthy controls and patients grouped by dataset. (B) Difference of extreme positive deviations between healthy controls and patients grouped by dataset (C) Difference of extreme negative deviations between healthy controls and patients grouped by dataset. (D) Difference of total extreme deviations between group. Patients are composed of a mixture of first episode (FDOPA\_01) and chronic psychosis (FDOPA\_02, FDOPA\_03), depending on the dataset. Asterisks indicates significance, \* indicates  $p < 0.05$ , \*\* indicates  $p < 0.01$  and \*\*\*\* indicates  $p < 0.001$ .*

*Figure 5. Clinical [<sup>18</sup>F]FDOPA Normative Model Treatment Response. ROC curves for the classification of treatment response (Responders vs Non-Responders/Resistant) using whole brain (solid lines) or striatal (dashed lines) individual average Z-score (A), whole brain positive extreme deviations (B), negative extreme deviations (C) and total extreme deviations (D) for the different clinical datasets. Dashed diagonal indicates the level of chance in all plots. All scores refer to [<sup>18</sup>F]FDOPA patient data after Combat harmonization.*

