Supplementary Information for

Neuromelanin-sensitive MRI as a non-invasive proxy measure of dopamine function in the human brain

Clifford M. Cassidy, Fabio A. Zucca, Ragy R. Girgis, Seth C. Baker, Jodi J. Weinstein, Madeleine E. Sharp, Chiara Bellei, Alice Valmadre, Nora Vanegas, Lawrence S. Kegeles, Gary Brucato, Un Jung Kang, David Sulzer, Luigi Zecca, Anissa Abi-Dargham, Guillermo Horga

Corresponding authors:

Guillermo Horga, MD, PhD
E-mail: HorgaG@nyspi.columbia.edu

Clifford Cassidy, PhD
E-mail: clifford.cassidy@theroyal.ca

This PDF file includes:

Supplementary text
Fig. S1 to S5
Table S1 to S5
References for SI reference citations

www.pnas.org/cgi/doi/10.1073/pnas.1807983116

Supplementary Information Text

Supplemental Methods

NM-MRI analysis: exclusion of voxels with few observations

To reduce the risk of type II error, voxels were excluded from the analysis if, after censoring of subject data points with missing or extreme values, the t-test of the regression coefficient β_1 for a particular analysis had fewer than 10 degrees of freedom (note that the degrees of freedom take into account the sample size with usable data in a given voxel as well as the number of model predictors). This voxel exclusion only applied to the analysis relating NM-MRI signal to dopamine release capacity given the smaller sample size of the PET dataset, and this analysis was thus performed for 1,341 resampled SN voxels (rather than for the full mask of 1,807 resampled voxels). Selecting exclusion thresholds anywhere between 8-11 degrees of freedom gave very similar results. See inset in **Fig. S4A** for distribution of degrees of freedom for all voxels in this analysis.

NM-MRI analysis: non-circular voxel selection for estimation of unbiased effect size

For voxelwise analyses, an unbiased measure of effect size was generated by using a leave-one-out procedure: for a given subject, voxels where the variable of interest was related to NM-MRI signal were first identified in an analysis including all subjects except for this (held-out) subject. The mean signal in the held-out subject was then calculated from this set of voxels. This procedure was repeated for all subjects so that each subject had an extracted, mean NM-MRI signal value obtained from an analysis that excluded them. This unbiased voxel selection and data extraction thus avoided statistical circularity. Unbiased estimates of effect size (Cohen's d or correlation coefficient) were then determined by relating these extracted NM-MRI signal values to variables of interest across held-out subjects and including the same covariates as in the voxelwise analysis and an additional covariate indexing subjects lacking full dorsal-SN coverage (due to dorsal-ventral gradient in NM-MRI signal intensity).

Post-mortem experiment

Post-mortem specimens of human midbrain tissue were obtained from The New York Brain Bank at Columbia University. Seven specimens were obtained, each from an individual who suffered from Alzheimer's disease or other non-PD dementia at the time of death (ages 44 to 90; for further clinical and demographic information see SI Appendix, Table S1). None suffered from PD, Parkinsonian syndromes, or any other movement disorder or neurodegenerative illness affecting the SN, based on neuropathological examination for accumulation of abnormal proteins such as alpha-synuclein, beta-amyloid or tau. One case showed marked decrease in neuronal density in the SN despite clearly identifiable NM. Analyses excluding this one case did not change the observed relationship between NM-MRI CNR and NM concentration (Results). Therefore, the data presented include this case to increase statistical power. Specimens were ~3-mm-thick slices of fresh frozen tissue from the rostral hemi-midbrain of the right hemisphere containing pigmented SN. They were stored at -80º C. These specimens were scanned using the NM-MRI protocol, after which they were dissected for analyses of NM tissue concentration. For the MRI scanning session, the specimens were progressively thawed to 20º C, as verified via a laser thermometer. Specimens were placed in a custom-made dish 3D-printed from MRI-compatible nylon polymer (NW Rapid Mfg, McMinnville, OR; see **Figs. 1A** and **1C**) and a matching grid-insert lid was placed on top of the specimen and affixed to hold the specimen in place. While secured

in the dish, specimens were fully immersed in an MRI-invisible lubricant (Fomblin® perfluoropolyether Y25; Solvay, Thorofare, NJ) and placed in a desiccator for 30 minutes to remove air from the tissue. Wells in the four cardinal points of the rim of the dish were filled with water to mark its location and orientation in the MRI images. The dishes were then placed on a custom stand inside a 32-channel, phased-array Nova head coil and scanned using the 2D GRE-MT NM-MRI sequence described above for *in vivo* imaging. The only changes in the post-mortem scanning protocol were an increase in the resolution (in-plane resolution=0.3125×0.3125 mm^2, slice thickness=0.60 mm) and a decrease in the FoV (160×80). After the scanning session, samples were refrozen in place and marked with gridlines by applying methylene blue dye (0.05% water solution [5 mg/10 ml]; Sigma-Aldrich, St. Louis, MO) to the tissue using the grid insert as a stamp. Guides built into the walls of the dish ensured that the orientation of the grid with respect to the specimen was fixed at all times. Within 4 days post-scanning, partially thawed specimens were dissected along gridlines after extensive removal of Fomblin® by dripping tissue slices, followed by gently rolling the surface of sections on ultraclean filter paper. Dissection and manipulation of tissue sections were performed with ceramic blades and titanium-and-plastic forceps to avoid contamination from iron. Each grid section (3.5 mm × 3.5 mm × ~3 mm, depending on the slice thickness), together with any adjacent partial grid sections, was weighed, stored separately in Eppendorf tubes, and frozen. Specimens were thus divided into 13-20 grid sections; the grid column and row number of each dissected grid section was coded.

Neurochemical measurement of NM concentration in post-mortem tissue: examination of chemical agents applied to post-mortem tissue

To test whether Fomblin® influenced NM measurement, small cubes of SN pars compacta with similar levels of pigmentation were dissected from a single healthy subject. Some cubes (n=3) were immersed in Fomblin®, then cleaned of the Fomblin® (drained and rolled on filter paper); the remaining cubes (n=5) were not immersed in Fomblin® as control samples. NM concentration was comparable in these two sets of cubes (mean ± standard deviation: 0.82 ± 0.08 versus 0.86 ± 0.09 µg NM/mg wet tissue, respectively; t_6=-0.62, p=0.56). The water-soluble methylene blue dye was efficiently removed during washing steps in our standard protocol to measure NM concentration; moreover, we confirmed that the absorption wavelength of this compound (with a peak near 680 nm) is far from that used in the determination of NM concentration (350 nm).

MRI measurement of NM signal in post-mortem tissue: automated removal of voxels showing edge artifacts and signal dropout

Processing of NM-MRI images included automated removal of low-signal voxels, including all voxels outside of the specimen or voxels within the specimen showing signal dropout. The threshold for exclusion of low-signal voxels was determined for each specimen based on the histogram of all voxels in the image, which was fitted using a kernel smoothing function. The threshold was defined as the signal corresponding to the minimum lying between the leftmost peak in the fitted histogram, corresponding to low-signal voxels outside of the specimen, and the rightmost peak, corresponding to higher-signal voxels within the specimen (consistent with a bimodal distribution).

To eliminate edge artifacts, the first step was to define the boundaries between the specimen and the surrounding space outside the specimen and between the specimen and areas of signal dropout. These boundaries were defined in 3D and 2D. To do so, boundary voxels of the specimen that lay directly next to low signal voxels (defined in the paragraph above) were labeled using the *bwperim* function in Matlab these boundary voxels were defined for the whole volume and also for a 2D flattened image created by

averaging over slices. These boundary voxels were removed from the specimen (first the 3D border voxels were removed from the 3D image, then the 2D boundary voxels, dilated by 2 voxels, were removed from the resulting flattened image). Finally, voxels with extreme signal values (Cook's distance>4/n in a constant-only linear regression model) relative to other voxels in the same 2D grid section were removed. The resultant 2D image, cleaned of edge artifacts, signal dropout and other outlier voxels, was carried forward to the final analysis steps.

Statistical Analysis of post-mortem data

A generalized linear mixed-effects (GLME) model including data across all grid sections g and specimens s was used to predict NM tissue concentration in each grid section based on mean NM-MRI CNR in the same grid section. GLME analyses used an isotropic covariance matrix and were fitted via maximum pseudo-likelihood estimation, as implemented via the Matlab function *fitglme*. Likelihood-ratio tests at $p<0.05$ favored reduced models without random slopes. Therefore, all models included random intercepts but not random slopes, as: $[NM]_{gs} = \beta_0 + \beta_1 \cdot \overline{\mathrm{CNR}}_{gs} + \sum_{i=2}^{n} \beta_i \cdot nuisance\ covariate_{gs} + b_{0s} + \varepsilon_{gs}$. The basic model only included mean NM-MRI CNR in a given grid section ($\overline{\mathrm{CNR}}_{gs}$) as a fixed-effects predictor. We observed that sections near the PAG tended to have relatively high signal intensity but low NM tissue concentration. Thus, an extended model included a binary variable for PAG presence in grid sections (PAG+, PAG-) and an interaction term of NM-MRI CNR × PAG as additional fixed-effects covariates (the interaction was significant at p=0.040, confirming that NM-MRI was less strongly related to NM concentration in PAG+ relative to PAG- regions). PAG+ grid sections (1 to 5 per specimen) were defined as those situated at the posterior-medial aspect of the specimen and consistent with the anatomical location of the PAG. Finally, a control analysis additionally included a fixed-effects covariate indicating the proportion of voxels containing SN for each grid section, defined as the proportion of voxels with CNR higher than 10% in grid sections deemed to contain SN upon visual inspection. This latter control analysis aimed to test whether regional variability in NM-MRI CNR would predict regional variability in NM tissue concentration even after accounting for changes in both measures as a mere function of the presence or absence of SN neurons in a given region (in combination with partial-volume effects).

PET imaging study: timing of post-amphetamine PET scan

Each subject received 2 post-amphetamine PET scans for the purposes of a separate experiment, which was previously published (1). This previous study aimed at assessing the time course of receptor internalization after an agonist challenge, measured via prolonged displacement of the D2 radiotracer [^{11}C]raclopride. PET scans were acquired in four sessions: baseline, 3 h after amphetamine, 5 to 7 h after amphetamine and 10 h after amphetamine. However, not all time points post-amphetamine were available for all subjects. Displacement was however highly stable and did not differ between the 3-h and the 5-7-h time points (1) (ΔBP_{ND} indeed strongly correlated across subjects between these two time points; r=0.75). For the purposes of the current study we only used one of these post-amphetamine scans: the one administered 5-7 hours post-amphetamine. We selected the 5-7-h time point, as we did in a prior report (2), because this was the time point with the most available data (missing for only 3/18 participants for whom it was substituted by data from the 3 hour time point). Displacement at 5 to 7 hours post-amphetamine –like displacement at 3 hours post-amphetamine– reflects the magnitude of dopamine release due to amphetamine, which is a combination of competition between dopamine and the radiotracer for binding to the receptor (3), and agonist-induced receptor internalization, both of which depend on the magnitude of agonist availability (4, 5). Thus, we believe that the 5-7-h time point was the optimal time point for this study due to the larger number of subjects with available data and given the

observed stability of the displacement between the 3-h and the 5-7-h time points. At the 10-h time point BP_{ND} tended to be higher, likely due to a decrease in receptor internalization following recycling of receptors. Examining the 11 subjects with PET data at 3 hours we found that the effect size of the correlation between NM-MRI CNR and ΔBP_{ND} at this 3 hour time point was similar to that at the 5-7 hour time point.

Arterial spin labelling (ASL) perfusion imaging study: CBF calculation

CBF was calculated with the following equation (6):

$$CBF = 6000 \times \lambda \frac{\left(1 - e^{\frac{ST}{T_{1t}}}\right) e^{\frac{PLD}{T_{1b}}}}{2\epsilon T_{1b}\left(1 - e^{-\frac{LT}{T_{1b}}}\right)} \left(\frac{PW}{SF_{PW} PR}\right)$$

Here, the longitudinal relaxation time (T1) of blood (T_{1b}) was assumed to be 1.6 s at 3.0T, T1 of tissue (T_{1t}) 1.2 s, partition coefficient (λ) 0.9, labeling efficiency (ϵ) 0.6, saturation time (ST) 2 s, labeling duration (LT) 1.5 s, and post-labeling delay (PLD) 1,525 ms. PW is the perfusion weighted or the raw difference image; PR is the partial saturation of the reference image, and SF_{PW} is an empirical scaling factor (32) used to increase the dynamic range of the PW.

Study Participants

Parkinson's disease study. Twenty-eight patients with idiopathic PD, as per UK Parkinson's Disease Society Brain Bank Criteria, were recruited either from the Center for Parkinson's Disease and other Movement Disorders at the Columbia University Medical Center or from the Michael J. Fox Foundation Trial Finder website. Patients were in the mild-to-moderate stages of disease (mean Unified Parkinson's Disease Rating Scale [UPDRS] off-medication score of 30, as administered by a movement disorders neurologist, with an average disease duration of 7.3 years; SI Appendix, Table S2). All patients had been receiving L-DOPA treatment for at least 6 months. Eleven of the 28 patients were scanned in the off-medication state (defined as more than 12 hours since the last dose of dopaminergic medication intake). NM-MRI signals did not differ between patients scanned on versus off medication. An age-matched sample of 12 healthy control participants (4 of which also participated in the psychosis study described below) was recruited from the local community (SI appendix, Table S2).

Schizophrenia sample. The inclusion criteria were: age 18-55 years; DSM-IV criteria for schizophrenia, schizophreniform or schizoaffective disorder as per the Structured Clinical Interview for DSM-IV Disorders (SCID-IV) (7, 8); negative urine toxicology; stable, outpatient medication-free status for at least three weeks. Exclusion criteria were: diagnosis of bipolar disorder, active substance use disorders (except tobacco use disorders) or current substance use based on urine toxicology. Patients were recruited from the outpatient research facilities at NYSPI. Psychosis severity was measured with the positive subscale of the Positive and Negative Syndrome Scale (PANSS; positive total score is referred to as PANSS-PT) (9); PANSS measures of negative symptoms and general psychopathology (PANSS-NT and PANSS-GT, respectively) were used as control variables.

CHR sample. CHR individuals were recruited from a longitudinal cohort study at the Center of Prevention and Evaluation (COPE) at NYSPI. COPE offers treatment to English-speaking individuals, aged 14 to 30 years, who are deemed to be at high-risk for psychosis. These CHR individuals were help-

seeking and met criteria for at least one of three psychosis-risk syndromes, as assessed with the Structured Interview for Prodromal Syndromes (SIPS) (10). This instrument was also used to measure the severity of attenuated positive psychotic symptoms (SIPS-PT); SIPS measures of negative symptoms (SIPS-NT) and general symptoms (SIPS-GT) were used as control variables.

Although our primary aim in this study was to assess correlates of psychosis severity within the clinical psychosis groups, we also recruited two separate, non-overlapping, groups of healthy controls for exploratory comparison purposes: one (n=30) age-matched to the schizophrenia group and another (n=15) age-matched to the CHR group. These groups were recruited through advertisements and word of mouth. Healthy controls were excluded for: current or past axis I disorder (except tobacco use disorder; as per the SCID-IV), history of neurological disorders or current major medical illness, and first degree relatives with a history of psychotic disorder.

See Table S2, S3, and S5 for demographic and clinical information for all relevant groups (for studies of Parkinson's disease, PET imaging, and psychosis respectively). Socio-economic status was measured with the Hollingshead interview (11).

Supplemental Results

Topographic relationships to dopamine release capacity and psychosis within SN voxels

SN voxels with a stronger relationship between NM-MRI CNR and dopamine release capacity tended to be more lateral and anterior, with no clear gradient along the superior-inferior axis (β_x=0.015, t_{1337} =5.87, $p=10^{-8}$; β_y=0.036, t_{1337}=17.1, $p=10^{-59}$; β_z=0.001, t_{1337} =-0.30, p=0.76; multiple-linear regression analysis predicting partial r across SN voxels as a function of their coordinates in x [absolute distance from the midline], y, and z directions).

Psychosis-overlap voxels, in turn, tended to predominate in ventral and anterior aspects of the SN, and to a lesser extent, in lateral aspects of the SN (β_x=0.22, t_{1803}=2.86, p=0.004; β_y=0.45, t_{1803}=6.14, $p=10^{-9}$; β_z=-0.65, t_{1803}=-6.69, $p=10^{-11}$; logistic regression analysis predicting presence of psychosis-overlap voxels as a function of their coordinates in x [absolute distance from the midline], y, and z directions).

Voxelwise analyses relating NM-MRI CNR to psychosis severity in each clinical group

When analyzing each clinical group separately, we found that higher CNR in the SN correlated significantly with more severe psychosis in schizophrenia (PANSS-PT scores: 404 of 1807 SN voxels; $p_{corrected}$=0.007, permutation test; peak voxel MNI coordinates [x, y, z]: 5, -22, -20 mm) and non-significantly with attenuated psychosis in CHR individuals (SIPS-PT scores: 116 voxels, $p_{corrected}$=0.26, permutation test; peak voxel MNI coordinates [x, y, z]: 10, -25, -18 mm; **Fig. 5**). Eliminating the CHR individual who was an outlier in the relationship of CNR in psychosis conjunction voxels to SIPS-PT (see **Fig. 5**) increased the number of voxels where higher CNR correlated to SIPS-PT, although this relationship did not reach statistical significance in the permutation test correcting for multiple comparisons (189 voxels, $p_{corrected}$=0.18).

Comparison of NM-MRI CNR across groups

Aged-matched healthy control groups (n=30 and n=15, respectively) did not significantly differ from patients with schizophrenia (n=33) or CHR individuals (n=25) in CNR of the psychosis-overlap voxels, although, numerically, the mean CNR was higher in schizophrenia than in age-matched controls and was higher in CHR individuals who went on to develop schizophrenia compared to those CHR individuals who did not and age-matched controls (**Fig. S5**). Nonetheless, part of the sample of CHR individuals is currently participating in follow-up procedures as part of a longitudinal study in the COPE clinic, so these results should be deemed as preliminary. Highly psychotic individuals (the top quartile 9/33, with PANSS-PT scores>19) had significantly higher NM-MRI CNR in psychosis voxels compared to controls (CNR = 17.1 ± 4.0 vs 12.8 ± 4.7; t_{37}=-2.10, p=0.043, Mann-Whitney p=0.044; **Fig. S5**). Consistent with many findings in individuals at risk of psychosis, this effect was in the same direction but did not yield significant results in the CHR individuals with most severe subsyndromal psychosis (the top quartile of individuals with the highest attenuated psychotic symptoms [n=7; SIPS-PT score>18] showed numerically higher CNR in psychosis voxels of the SN compared to controls; t_{20}=-0.79, p=0.44).

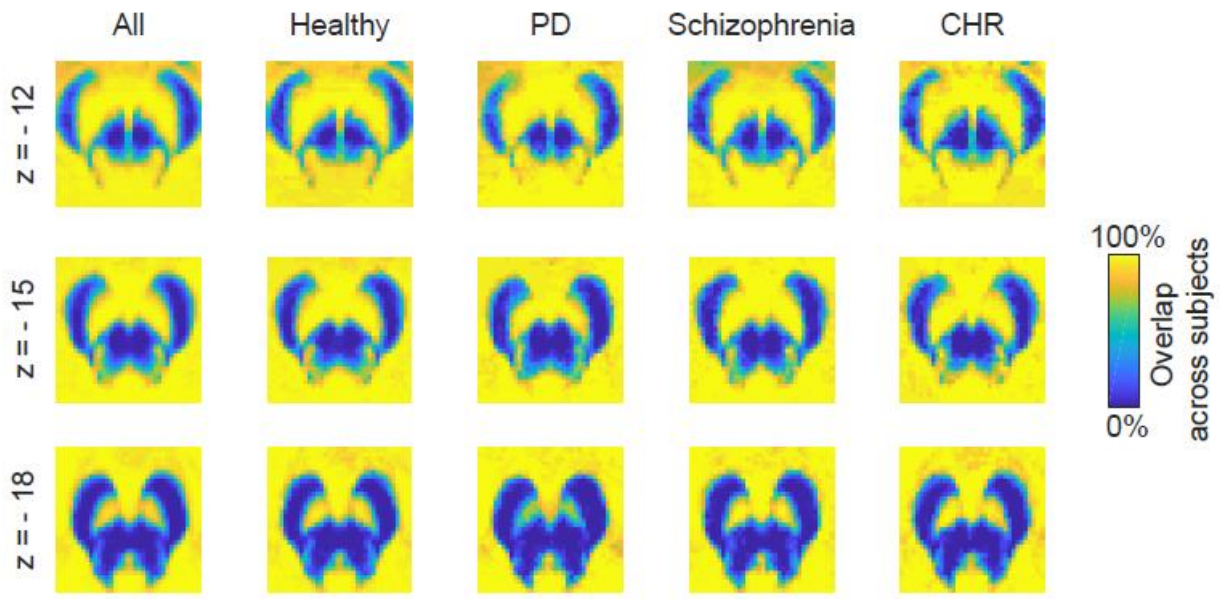

Figure S1. Quality check of spatial normalization procedures for all study groups. Overlap images indicate the percentage of subjects with spatially overlapping signal in SN and outside the midbrain for superior (z=-12), middle (z=-15), and inferior (z=-18) slices (row, from top to bottom), by group (column). These images were made by creating binary maps of each subject's preprocessed NM-MRI image (thresholded at CNR=10%) and calculating the percent overlap for each voxel across all subjects in each specific group. PD: Parkinson's disease. CHR: clinical high-risk individuals.

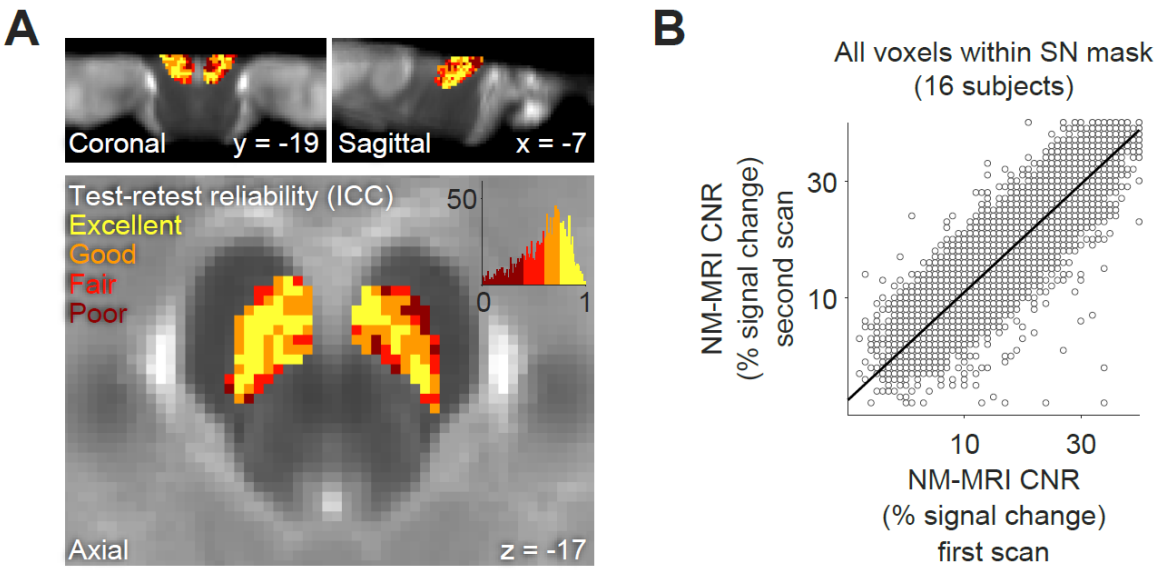

Figure S2. Test-retest reliability of NM-MRI across SN voxels. (A) Map of intraclass correlation coefficient (ICC) values across voxels in the SN, derived from 2 scans obtained approximately 1 hour apart in the same day (each of the 2 scans was obtained in 16 subjects). Two-way mixed, single score ICC value was calculated (ICC(3,1); (12)) for each voxel. This ICC score reflects consistency across the first and second scans. Standard thresholds are used for interpretability of ICC values: “excellent” reliability for ICC over 0.75, “good” reliability for ICC between 0.75 and 0.6, “fair” reliability for ICC between 0.6 and 0.4, and “poor” reliability for ICC under 0.4 (13). The inset histogram shows distribution of ICC (x axis) values across all in-mask SN voxels (y axis indicates voxel count). The median ICC across voxels was 0.64 (0.35 interquartile range). The ICC for NM-MRI averaged in the whole ICC mask was ‘excellent’ (ICC = 0.95). Calculating absolute agreement across scans for each voxel yielded very similar values (the median ICC(2,1) was 0.63, 0.35 interquartile range). Note that voxels with poor reliability tend to lie around the edge of the SN mask, which by definition may include voxels outside of SN proper in some subjects. (B) Scatterplot showing agreement in NM-MRI CNR for all voxels and all subjects between the 2 scans. This sample consisting of 8 healthy individuals and 8 patients with schizophrenia, with mean age 33.8 ± 13.3 years (a subset of the participants in the current study), was collected as part of a separate study.

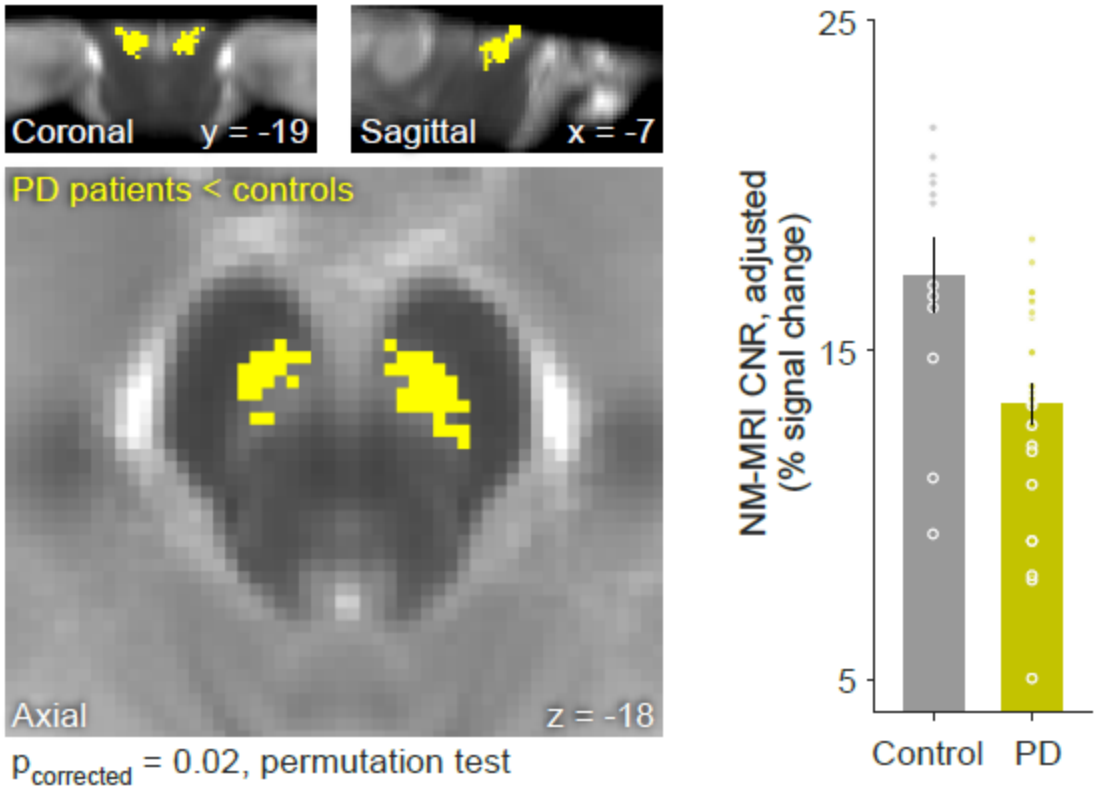

Figure S3. Comparison of PD patients and matched controls. The map on the left shows SN voxels where PD patients had decreased NM-MRI CNR (yellow voxels; thresholded at p<0.05, voxel-level), overlaid on the NM-MRI template image. The combined scatterplot and bar plot on the right shows mean NM-MRI CNR values extracted from the yellow voxels plotted by diagnostic group (PD patients versus sample of matched, healthy controls) for visualization purposes. Each data point is one subject. Error bars are means and SEM. The Cohen's d effect size for the group difference in the plotted data was d=1.08, but note that this estimate is biased due to circular voxel selection; the unbiased effect-size was d=0.89 as calculated via a leave-one-out procedure for unbiased voxel selection (see NM-MRI analysis: non-circular voxel selection for estimation of unbiased effect size, SI Appendix).

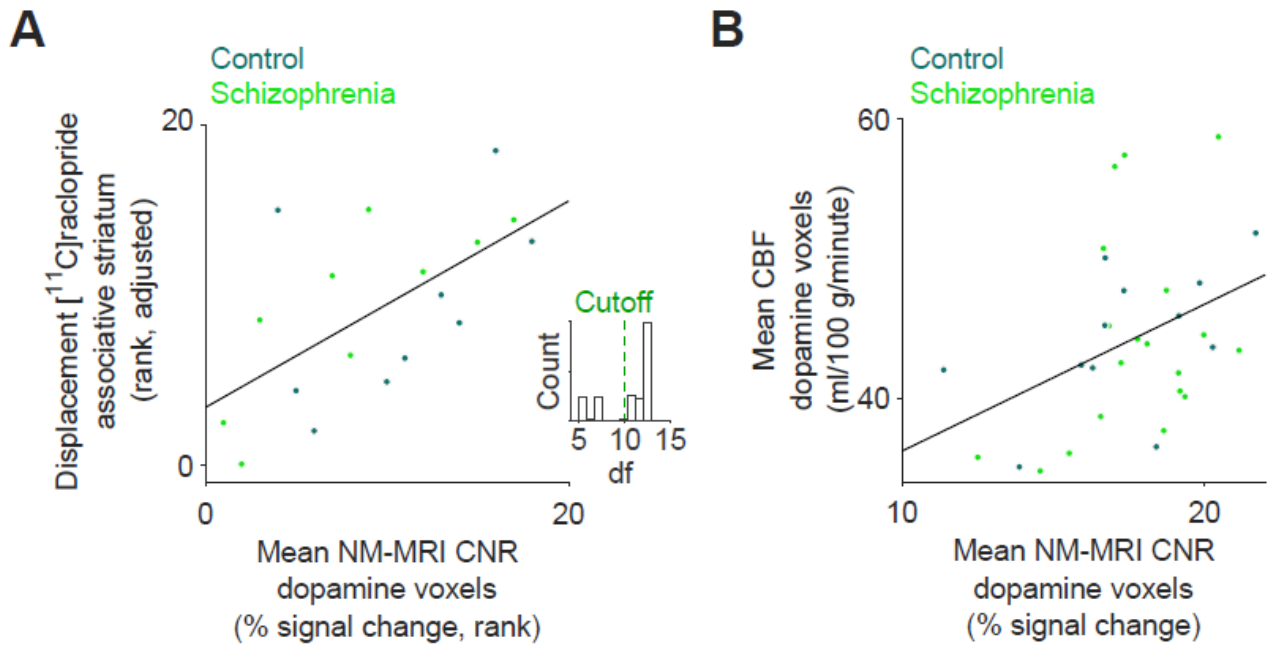

Figure S4. NM-MRI CNR correlates with measures of dopamine function across individuals without neurodegenerative illness. The same scatterplots from Fig. 4A and B (main text) are shown here with the participant groups shown in different colors (controls are grayish dark green; schizophrenia patients are light green). No interactions by group were found for either analysis (all $p>0.3$). When controlling for previous history of exposure to antipsychotic medication, the correlation of dopamine release capacity to NM-MRI CNR extracted from dopamine voxels remained significant ($\rho=0.55$, $p=0.049$) arguing against the possibility that previous treatment could account for this relationship. The inset histogram in A shows the distribution of degrees of freedom (df; for t-test of the regression coefficient β_1) for all analyzed voxels in the voxelwise analysis relating dopamine release capacity to NM-MRI CNR. The presence of some NM-MRI scans lacking full-coverage of the SN (due to interindividual differences in anatomy) led to reduction in degrees of freedom for some (more dorsal) voxels. These voxels are represented as the minor mode on the left of the histogram, where all voxels have degrees of freedom below 10. Thus, the cutoff for voxel exclusion was set at $df<10$ (green broken line on the histogram).

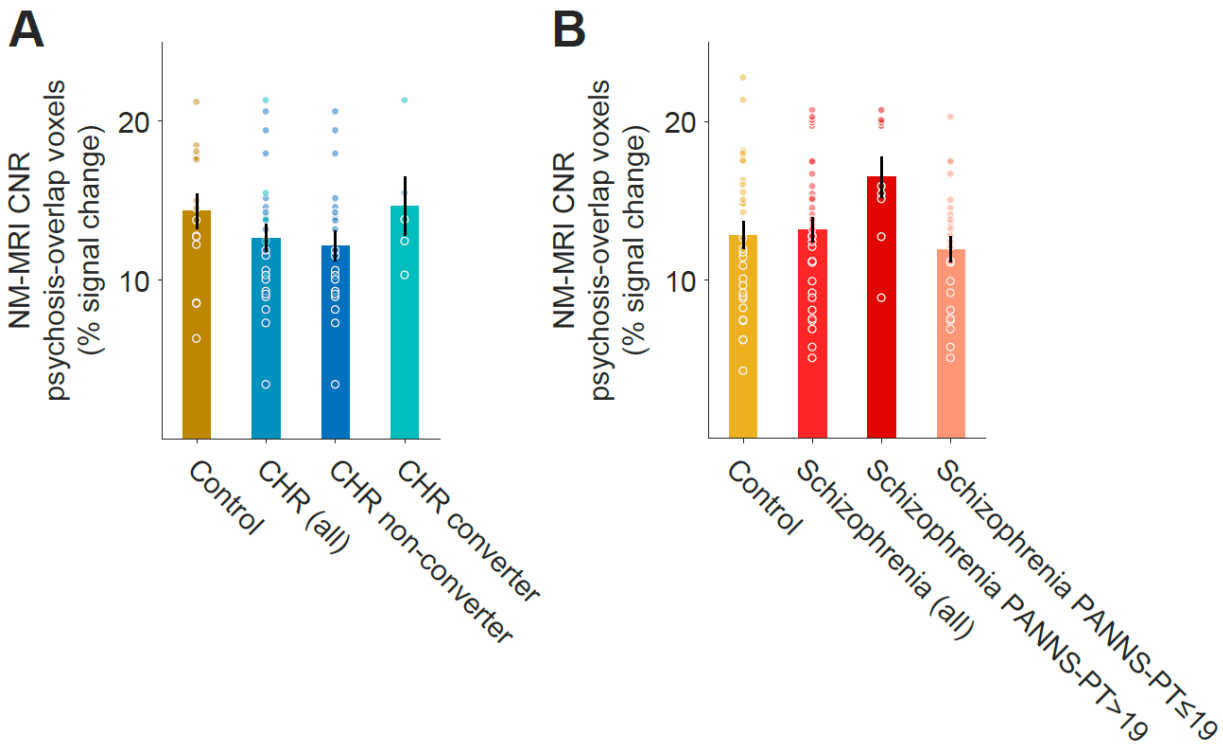

Figure S5. Comparison of healthy-control groups and clinical groups on NM-MRI CNR extracted from psychosis-overlap voxels (pink voxels in Fig. 5, main text). (A) Comparison of clinical high-risk (CHR) individuals for psychosis to age-matched healthy controls. All CHR individuals are shown (blue bar, second from left) as well as the subgroups of CHR individuals who did and did not subsequently convert to full-blown psychotic illness (two rightmost bars in different blue shades). (B) Comparison of unmedicated patients with schizophrenia to age-matched healthy controls (note that this group represents a separate cohort of controls from that shown in A). All schizophrenia patients are shown (red bar, second from left) as well as the subgroups of schizophrenia patients with high positive symptom scores (darker red; PANSS-PT scores>19, top quartile, n=9) and low positive symptom scores (lighted red; PANSS≤19, n=24). Highly symptomatic schizophrenia patients showed higher NM-MRI CNR extracted from psychosis-overlap voxels compared to healthy controls (t_{37}=-2.10, p=0.043) and low symptom patients (t_{31}=-2.93, p=0.0064). No statistically significant differences were seen in other group comparisons. Error bars depict means and SEM. Individual data points represent subjects. Note that this is not representative of all SN voxels or voxels that may have shown trends for diagnostic effects (but that did not survive the corrected threshold). The lack of diagnostic effects in psychosis-overlap voxels simply suggests that such effects are not present in the SN regions exhibiting psychosis effects.

Table S1. Clinical and demographic information for post-mortem specimens

Sample ID	Age	Sex	Diagnosis	Cold post-mortem interval (hours)	Frozen post-mortem interval (hours)
1	81	F	AD	6.5	19.3
2	76	F	AD	5.5	18.2
3	44	M	AD	2.8	20.8
4	72	M	AD	-	9.0
5	>88	-	AD (probable)	3.1	25.7
6	>88	F	Dementia	0.8	16.5
7	84	F	AD (possible)	1.8	23.7

AD: Alzheimer's disease. F: female; M: male.

Table S2. Clinical and demographic characteristics of clinical samples for Parkinson's disease study.

Characteristic		Parkinson's disease (n=28)	Matched healthy controls (n=12)	PD vs. matched controls (p-value)
Sociodemographic characteristics				
Age, years [range]		63 ± 1.1 [51-72]	62 ± 2.44 [51-73]	0.64
Sex (male)		24 (86%)	9 (75%)	0.41
Race/ ethnicity	African-American	0	2	0.014
	Asian	0	2	
	Caucasian	26	7	
	Hispanic	2	1	
	Mixed	0	0	
Clinical characteristics				
UPDRS	Off medication	30 ± 1.8	1.5 ± 0.43	<0.001
	On medication	20 ± 1.5	-	-
MoCA		27.4 ± 0.43	29 ± 0.34	0.18
Disease duration, years		7.3 ± 0.64	-	-

Means ± standard error are given for continuous variables; number (and percentage) are given for categorical variables. P-values for group comparison of Parkinson's disease patients and healthy controls are given based on two-sample t-tests for continuous variables and X^2 tests for categorical variables. Parental SES: parental socio-economic status as measured via the Hollingshead scale. UPDRS: Unified Parkinson's Disease Rating Scale. MoCA: Montreal Cognitive Assessment.

Table S3. Sociodemographic and Positron Emission Tomography (PET) data for PET study sample

Characteristic		Healthy controls (n=9)	Patients with schizophrenia (n=9)
Sociodemographic characteristics			
Sex (male)		6 (66%)	6 (66%)
Age, years [range]		31.9 ± 2.5 [23-59]	31.9 ± 3.2 [20-47]
Race/ethnicity	African-American	4	5
	Asian	1	1
	Caucasian	2	2
	Hispanic	2	1
Subject SES		45.1 ± 5.1	20.7 ± 2.8
Parental SES		47.9 ± 2.0	36.3 ± 4.3
Tobacco Users		2 (22%)	3 (33%)
Antipsychotic history (drug naïve/drug free[1])		-	5/4
PET parameters			
Hours post-amphetamine [2]		5.3 ± 1.3	5.3 ± 1.0
Amphetamine level (ng/mL) [2]		74.1 ± 13.3	70.1 ± 10.4
Radiotracer Mass (μg) [3]		2.3 ± 2.1, 2.5 ± 1.3	1.8 ± 0.6, 2.0 ± 0.9
Radiotracer Dose (mCi) [3]		8.4 ± 2.8, 11.2 ± 2.6	10.0 ± 3.1, 10.8 ± 2.9

Means ± standard error are given for continuous variables; number (and percentage) are given for categorical variables. P-values for group comparisons are given based on two-sample t-tests for continuous variables and X^2 tests for categorical variables. [1]Antipsychotic medication status was considered "drug-naïve" if lifetime exposure <6 weeks and none in past 3 weeks, and "drug-free" if none in past 3 weeks. [2]Mean at start of scan 5-7 hours post-amphetamine. [3]Mean for scans at baseline and 5-7 hours post-amphetamine. SES: socioeconomic status.

Table S4. Voxelwise correlations of NM-MRI signal to dopamine release capacity in different striatal subdivisions

Striatal division	SN voxels where NM-MRI correlated to increased dopamine release capacity (count)	Corrected p-value
Whole striatum	217	0.046
Associative striatum	225	0.042
Sensory-motor striatum	128	0.15
Ventral striatum	305	0.013

Table S5. Characteristics of psychosis sample and specific healthy-control groups

Characteristic		Schizophrenia (n=33)	Controls for schizophrenia (n=30)	Clinical high-risk individuals (n=25)	Controls for CHR (n=15)	Schizophrenia vs. controls (p-value)	CHR vs. controls (p-value)
Socio-demographic characteristics							
Age, years		33.9 ± 2.2	34 ± 2.2	25.6 ± 0.97	23.5 ± 0.86	0.97	0.15
Sex (male)		23 (70%)	18 (60%)	13 (52%)	9 (60%)	0.42	0.46
Race/ ethnicity	African-American	17 (52%)	12 (40%)	6 (24%)	3 (20%)	0.71	0.28
	Asian	2 (6%)	1 (3%)	1 (4%)	2 (13%)		
	Caucasian	7 (21%)	6 (20%)	10 (40%)	3 (20%)		
	Hispanic	5 (15%)	7 (23%)	2 (8%)	4 (27%)		
	Mixed	2 (6%)	4 (13%)	6 (24%)	2 (13%)		
Parental SES		41.7 ± 2.4	41.9 ± 2.2	44.5 ± 2.2	56.9 ± 3.8	0.95	0.008
Clinical Characteristics							
Antipsychotic history (drug naïve/drug free)[1]		17/16	-	-	-	-	-
Nicotine users		3 (11%)	3 (12%)	1 (5%)	0 (0%)	0.96	1
PANSS	Positive total [range, 7-49]	16.9 ± 0.93	7.1 ± 0.09	-	-	<0.0001	-
	Negative total [range, 7-49]	15.4 ± 0.99	8.9 ± 0.5	-	-	<0.0001	-
	General total [range, 16-112]	32.8 ± 1.8	16.7 ± 0.39	-	-	<0.0001	-
SIPS	Positive total [range, 0-30]	-	-	15.6 ± 0.70	2.8 ± 0.65	-	<0.0001
	Negative total [range, 0-36]	-	-	17.4 ± 0.91	2.7 ± 1.02	-	<0.0001
	Disorganization total [range, 0-24]	-	-	11.2 ± 0.53	1.5 ± 0.35	-	<0.0001
	General total [range, 0-24]	-	-	13.7 ± 0.68	2 ± 0.53	-	<0.0001

Means ± standard error are given for continuous variables; number (and percentage) are given for categorical variables. [1]Antipsychotic medication status was considered "drug-naïve" if lifetime exposure <6 weeks and none in past 3 weeks, and "drug-free" if none in past 3 weeks. P-values for group comparisons are given based on two-sample t-tests for continuous variables and X^2 tests for categorical variables. SES: socio-economic status. PANSS: Positive and Negative Syndrome Scale (positive or psychotic symptoms of schizophrenia include hallucinations and delusions; negative symptoms include emotional withdrawal and amotivation). SIPS: Structured Interview for Prodromal Syndromes.

References

1. Weinstein JJ*, et al.* (2017) PET imaging of dopamine-D2 receptor internalization in schizophrenia. *Mol Psychiatry*.
2. Cassidy CM*, et al.* (2018) A Perceptual Inference Mechanism for Hallucinations Linked to Striatal Dopamine. *Curr Biol* 28(4):503-514 e504.
3. Skinbjerg M*, et al.* (2010) D2 dopamine receptor internalization prolongs the decrease of radioligand binding after amphetamine: a PET study in a receptor internalization-deficient mouse model. *Neuroimage* 50(4):1402-1407.
4. Guo N*, et al.* (2010) Impact of D2 receptor internalization on binding affinity of neuroimaging radiotracers. *Neuropsychopharmacology* 35(3):806-817.
5. Ito K, Haga T, Lameh J, & Sadee W (1999) Sequestration of dopamine D2 receptors depends on coexpression of G-protein-coupled receptor kinases 2 or 5. *Eur J Biochem* 260(1):112-119.
6. Wong AM, Yan FX, & Liu HL (2014) Comparison of three-dimensional pseudo-continuous arterial spin labeling perfusion imaging with gradient-echo and spin-echo dynamic susceptibility contrast MRI. *J Magn Reson Imaging* 39(2):427-433.
7. First M, Spitzer R, Gibbon M, & Williams J (1995) *Structured Clinical Interview for DSM-IV Axis I Disorders (SCID-I/P, Version 2.0)* (Biometrics Research Dept., New York State Psychiatric Institute, New York).
8. Nurnberger JI, Jr.*, et al.* (1994) Diagnostic interview for genetic studies. Rationale, unique features, and training. NIMH Genetics Initiative. *Arch Gen Psychiatry* 51(11):849-859; discussion 863-844.
9. Kay SR, Fiszbein A, & Opler LA (1987) The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophr Bull* 13(2):261-276.
10. Miller TJ*, et al.* (2003) Prodromal assessment with the structured interview for prodromal syndromes and the scale of prodromal symptoms: predictive validity, interrater reliability, and training to reliability. *Schizophr Bull* 29(4):703-715.
11. Hollingshead AB (1975) *Four factor index of social status.* (Working paper published by the author., New Haven, Connecticut).
12. Shrout PE & Fleiss JL (1979) Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 86(2):420-428.
13. Cicchetti DV (1994) Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*.6(4):pp.