






# A single-cell atlas of the human substantia nigra reveals cell-specific pathways associated with neurological disorders

Devika Agarwal<sup>1,2,8</sup>, Cynthia Sandor<sup>3,8</sup>, Viola Volpato<sup>3,8</sup>, Tara M. Caffrey <sup>2,6</sup>, Jimena Monzón-Sandoval<sup>3</sup>, Rory Bowden <sup>4</sup>, Javier Alegre-Abarrategui <sup>2,5,7</sup>, Richard Wade-Martins <sup>2</sup> & Caleb Webber <sup>2,3</sup>✉

We describe a human single-nuclei transcriptomic atlas for the substantia nigra (SN), generated by sequencing approximately 17,000 nuclei from matched cortical and SN samples. We show that the common genetic risk for Parkinson's disease (PD) is associated with dopaminergic neuron (DaN)-specific gene expression, including mitochondrial functioning, protein folding and ubiquitination pathways. We identify a distinct cell type association between PD risk and oligodendrocyte-specific gene expression. Unlike Alzheimer's disease (AD), we find no association between PD risk and microglia or astrocytes, suggesting that neuroinflammation plays a less causal role in PD than AD. Beyond PD, we find associations between SN DaNs and GABAergic neuron gene expression and multiple neuropsychiatric disorders. Conditional analysis reveals that distinct neuropsychiatric disorders associate with distinct sets of neuron-specific genes but converge onto shared loci within oligodendrocytes and oligodendrocyte precursors. This atlas guides our aetiological understanding by associating SN cell type expression profiles with specific disease risk.

<sup>1</sup> Weatherall Institute of Molecular Medicine, John Radcliffe Hospital, University of Oxford, Oxford, UK. <sup>2</sup> Oxford Parkinson's Disease Centre, Department of Physiology, Anatomy, Genetics, University of Oxford, Oxford, UK. <sup>3</sup> UK Dementia Research Institute, Cardiff University, Cardiff, UK. <sup>4</sup> Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford OX2 7BN, UK. <sup>5</sup> Department of Neuropathology, University of Oxford, Oxford, UK. <sup>6</sup> Present address: Djavad Mowafaghian Centre for Brain Health, University of British Columbia, Vancouver, BC, Canada. <sup>7</sup> Present address: Division of Brain Sciences, Imperial College London, London, UK. <sup>8</sup> These authors contributed equally: Devika Agarwal, Cynthia Sandor, Viola Volpato. ✉email: [webberc4@cardiff.ac.uk](mailto:webberc4@cardiff.ac.uk)

The identification of the cell types relevant to a given disease is key to understanding the causal processes underlying the molecular aetiology. Many disorders, including neurological and psychiatric, are strongly influenced by genetic variation altering gene function. Therefore, by associating genetic variation with particular genes, those genes' spatial expression patterns can then implicate specific cell types with that disorder. The cell type (s) associated with the genetic risk for a disease are not necessarily those cell types most directly associated with the defining symptoms. For Alzheimer's disease (AD), intersecting risk-associated common genetic variation with cell-type-specific gene expression proposed that the genetic risk for AD most significantly influences microglia, moving the research focus away from the neurons whose loss underlies symptoms, towards a role for neuroinflammation in this neuronal loss<sup>1</sup>. It is thus critical to re-examine our aetiological assumptions when the data to do so are available.

The main hallmark of Parkinson's disease (PD), the most common progressive neurodegenerative movement disorder, is the selective loss of dopaminergic neurons (DaNs) in the substantia nigra (SN). Midbrain DaNs from the SN have a key role in the regulation of movement, cognition, motivation and reward, and their loss underlies deficits in fine motor control observed in PD<sup>2</sup>. However, while the loss of DaNs underlies key PD pathology, roles for other cell types such as astrocytes and microglia have been proposed that contribute to this loss<sup>3,4</sup>. Unfortunately, the human SN is sorely understudied compared with the neocortex, for which single-cell/nuclei studies<sup>5,6</sup> have comprehensively mapped and characterised the diverse cell-type populations and identified cell-type-specific disease associations<sup>6,7</sup>. A comparable systematic and unbiased survey of cell-type-specific gene expression across the human SN will help us to identify the role of neuroglia in the selective vulnerability of the DaNs in PD, and provide significant insight into the potential contributions of the SN cell types to other neurological disorders.

In this study, single-nuclei RNA sequencing was performed on the human SN and cortex regions from the same individuals to resolve their regional cellular diversity and to identify region-specific neuronal and non-neuronal cell-type differences. From this cellular atlas, we mapped common risk variants for a range of brain-related disorders/traits to specific cell types. We performed a cell-type-specific gene network analysis of PD and other psychiatric risk genes in the SN. Among many multiple disease-risk/cell-type association, we show that genetic risk in PD is indeed associated with DaNs, but also with oligodendrocytes (ODCs), giving new insights into the causes of PD and other disorders. By contrast, we find no association between PD risk and microglia, suggesting that neuroinflammation is not a key causal process in PD.

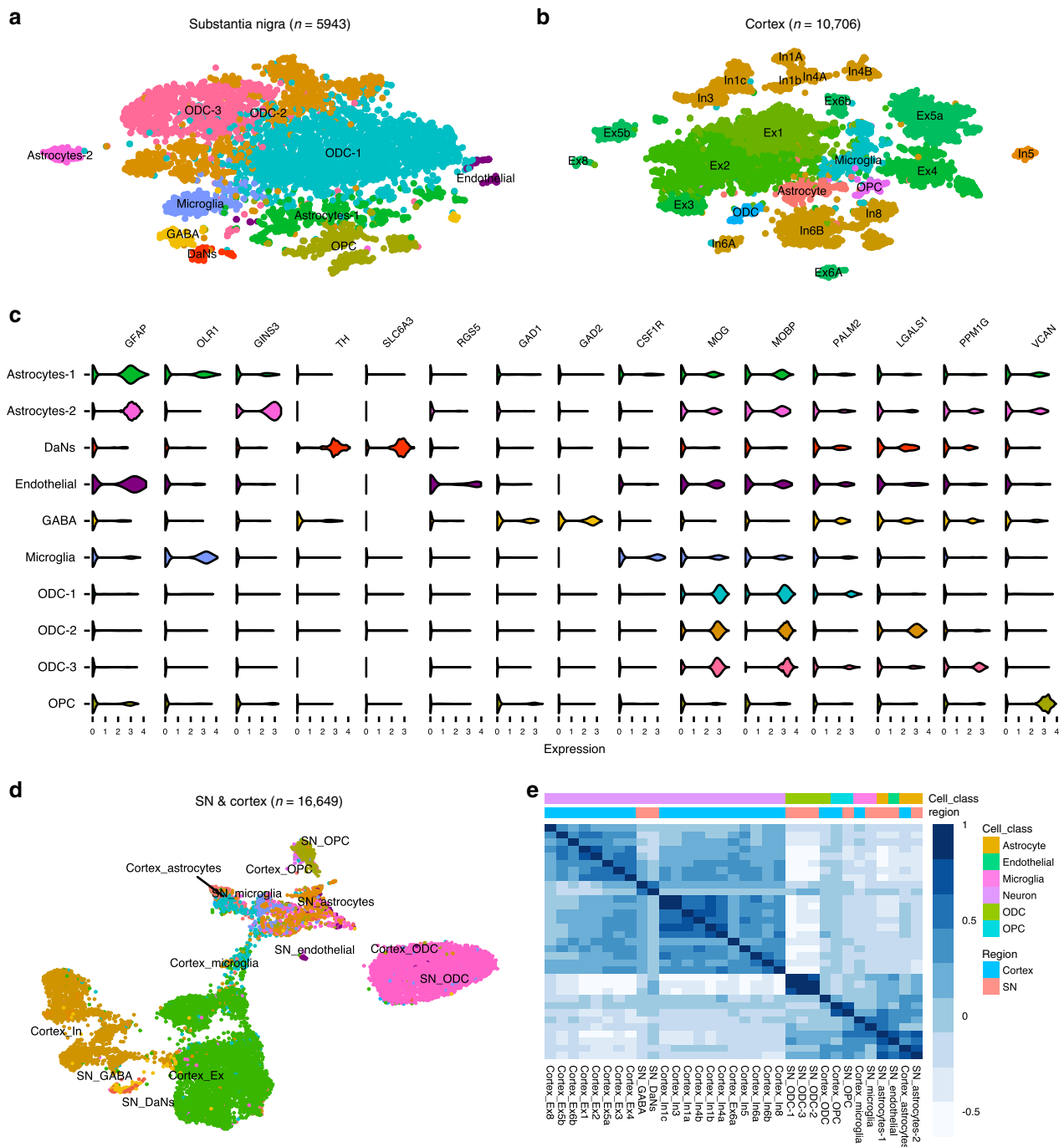
## Results

**Single-nuclei sequencing of the human SN and cortex.** We sequenced the transcriptomic profiles of 10,706 nuclei and 5943 nuclei from the cortex (middle frontal gyrus) and SN, respectively, of 12 matched samples (including two SN replicates) from five human postmortem brains using the 10x Genomics Chromium platform (Supplementary Table 1, Supplementary Data 1, "Methods"). We identified ten distinct cell populations across all samples within the SN (Fig. 1a, Supplementary Fig. 1, Supplementary Note 1), which included (i) astrocytes (*GFAP*) with two subtypes: astrocyte-1 population expressing neuro-inflammatory genes (*OLR1*) and an astrocyte-2 (*GINS3*) population expressing genes associated with growth and reparative functions (Supplementary Table 2 and Supplementary Fig. 2)<sup>8</sup>, (ii) oligodendrocytes (ODCs) (*MOG*, *MOBP*) with

three subtypes discriminated by oligodendrocyte marker genes *PALM2*, *LGALS1* & *PPM1G*, (iii) endothelial cells (*RGS5*), (iv) microglia cells (*CSF1R*), (v) oligodendrocyte precursor cells (OPCs) (*VCAN*), (vi) DaNs (*TH* and *SLC6A3*), neuronal population of the SN pars compacta (Supplementary Fig. 2) and (vii) GABAergic neurons, neuronal population of the SN pars reticulata expressing gamma-aminobutyric acid (GABA) receptors *GABRA1* and *GABRB2* and the enzymes *GAD1* and *GAD2* required for GABA neurotransmitter synthesis (Fig. 1a, c, Supplementary Data 2 and 3, Supplementary Fig. 2). In the cortex, we identified six distinct cell populations including: astrocytes, excitatory neurons (Ex), inhibitory neurons (In), ODCs, OPCs and microglia, but no distinct cluster for endothelial cells (Fig. 1b, Supplementary Figs. 3 and 4, Supplementary Data 2 and 3, Supplementary Note 1). A joint clustering of both SN and cortex regions revealed that the cortex and SN form distinct clusters by cell types (neuronal cells; ODC; astrocyte, microglia, OPC and endothelial cells) rather than by region (Fig. 1d, e, Supplementary Note 1).

We captured a higher proportion of nuclei from glia in the SN (95.5% glia) represented mainly by ODCs (72%) (Supplementary Table 3a) than that obtained from the cortex (12% glia) (Supplementary Table 3b). As our cortex and SN atlases have been generated from the same individuals through the same process, this difference in glial proportion likely reflects genuine variation in the cellular composition of different brain regions. The observed proportions of different cell populations in our SN atlas were consistent with histopathological studies, reporting ODCs as the most frequent glial cell population (45–75% across all brain regions, 62% for SN)<sup>9</sup> and recent bioinformatics predictions of tissue cell-type composition<sup>10</sup>. Despite significant variation in the numbers of cell types captured between inter- and intra-individual samples (Supplementary Table 3b), we observed consistent clustering by cell type between replicates, across samples and regions (Supplementary Fig. 5), suggesting we have repeatedly captured the same resident cell populations.

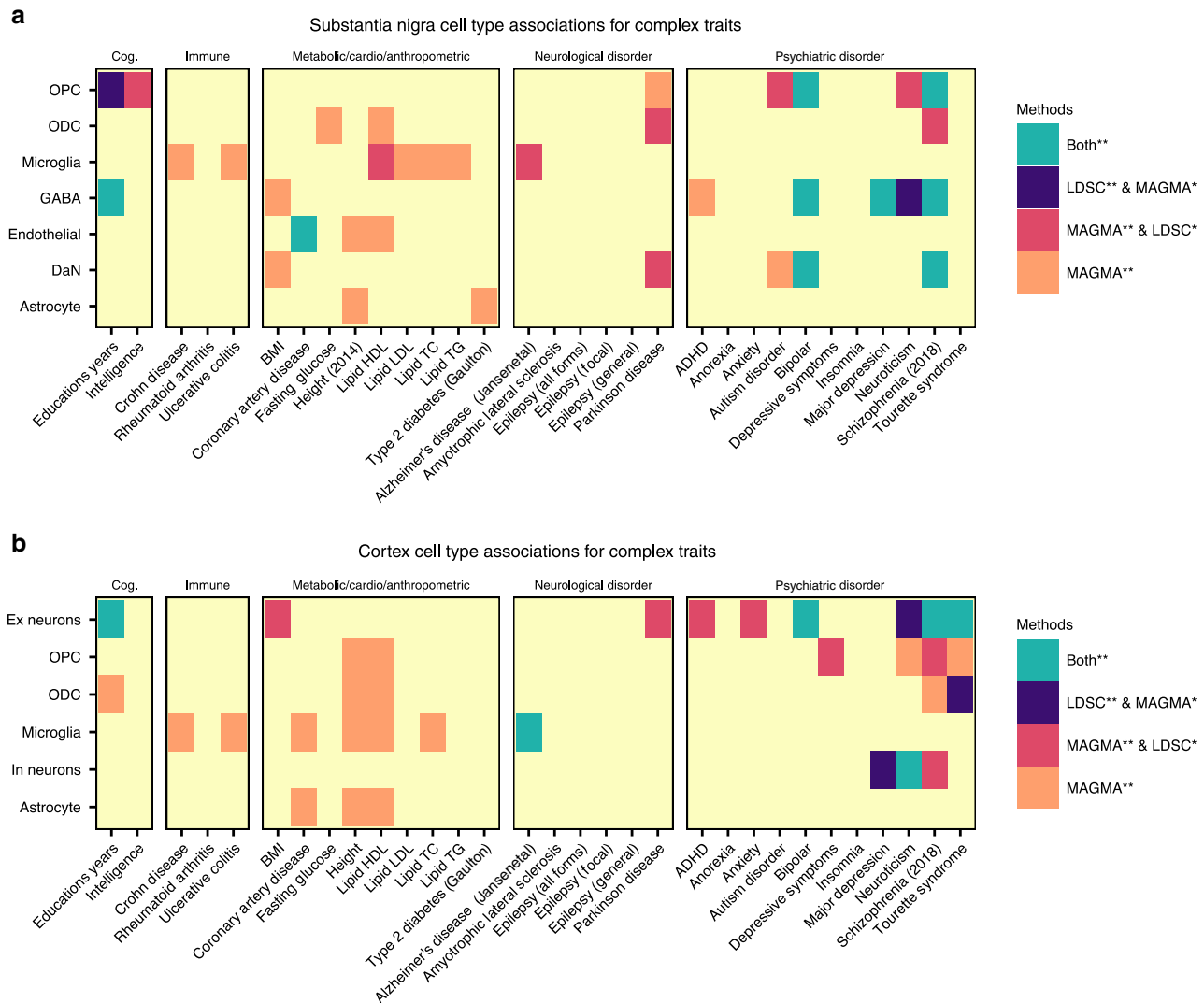
**Complex trait genetics highlights distinct brain cell types.** We used the cell-type-specific gene expression patterns from the human SN transcriptomic cellular atlas to identify specific SN cell types through which genetic variants contributing to each of 30 human complex traits (Supplementary Table 4) might be acting, employing two distinct methods (LD score regression (LDSC)<sup>11</sup> and multi-marker analysis of genomic annotation (MAGMA)<sup>12</sup>) (Supplementary Table 5). We observed for the first time in human a significant association between PD genetic risk and genes with DaN-specific expression patterns ( $q_{\text{MAGMA}} = 4.6 \times 10^{-3}$ , Fig. 2a), supported by recent observations made using mouse expression patterns of predicted PD GWA-risk genes<sup>13,14</sup> and contrasting studies proposing that PD risk is not associated with neurons<sup>15</sup>. We also identified a second association between PD genetic risk and genes with ODC-specific expression patterns ( $q_{\text{MAGMA}} = 0.035$ ; Fig. 2a). To evaluate whether the fraction of PD genetic risk associated with the DaN or OPC cell types was overlapping or distinct, we performed conditional analyses by running LDSC with either the ODC or the DaN gene set as the control: we found that the fraction of PD genetic risk contributing to the ODC association was distinct to that fraction associated with DaNs (Supplementary Table 6,  $p$  value associated with an LDSC coefficient  $< 0.05$  (here,  $p = 7 \times 10^{-3}$ )), proposing distinct PD-associated cell etiologies within the SN. In the cortex map, we found a significant Ex neuron cell association with PD genetic risk variants ( $q_{\text{MAGMA}} = 6.1 \times 10^{-3}$ , Fig. 2b, Supplementary Table 7). Our cortex/SN paired tissue sample study design enabled conditional analyses without effects due to individual/



**Fig. 1** Single-nuclei transcriptomic cell atlas of the adult human cortex and substantia nigra (SN). T-distributed stochastic neighbour embedding (t-SNE) plot of gene expression relationships amongst the single-nuclei cells from five individuals in the (a) substantia nigra (SN) ( $n = 5943$ ) and (b) cortex ( $n = 10,706$ ). c Violin plots of expression values ( $\log_{10}$  TPM values) of enriched cell-type-specific markers for the cell types in the SN (Supplementary Data 3). d Uniform manifold approximation and projection (UMAP) plot of both the cortex and substantia nigra cell types ( $n = 16,649$  cells) showing distinct clustering by cell type. e Correlation heatmaps showing hierarchical clustering of Pearson correlation scores calculated between averaged cell-type subclusters in both regions. The transcriptional correlation is largely explained by cell type and not by the region of origin. DaNs dopaminergic neurons, Ex excitatory neurons, GABA GABAergic neurons, In inhibitory neurons, ODC oligodendrocytes, OPC oligo-precursor cells.

study variation to be performed for cell types across these two brain regions (Supplementary Note 3). We examined the effects of the differing neuron/glia cell proportions between the SN and cortex upon genetic risk/cell associations by creating artificially matched SN/cortex cellular atlases possessing the same proportion of glial cells by randomly sampling the original cellular populations (“Methods”). We observed a large agreement in the genetic risk/cell associations between the original and the

homogenous cell atlas (Supplementary Fig. 6) (SN  $R = 0.83$  ( $p < 10^{-16}$ ) and cortex  $R = 0.72$  ( $p < 10^{-16}$ )), showing that the cellular proportions do not obscure the cell-type associations within the same tissue. Nevertheless, only the cross-tissue cell-type conditional analysis conducted within the homogeneous cell atlas revealed that the PD genetic risk associated with the SN DaNs and cortex Ex gene expression profiles was indistinguishable (Supplementary Table 8,  $p = 0.15$  (homogenous cell atlas)/ $p = 0.049$



**Fig. 2 Identification of brain cell types associated with diverse human complex traits.** We used two approaches to identify the associations between genetic risk variants of different complex trait and cell types from (a) substantia nigra and (b) cortex: stratified LD score regression (LDSC) ( $p$  value associated with an LDSC Coefficient (“Methods”)) and the MAGMA gene set analysis (one-sided positive two-sample  $t$ -test). The heatmap colours give different degrees of significance with both methods or either method, an asterisk (\*) and double asterisks (\*\*) indicate nominally significant  $p$  value ( $<0.05$ ) and  $q$  value (Bonferroni correction for the number of cell types tested). The different traits were clustered by category: cognitive phenotypes (Cog.), autoimmune diseases (Immune), metabolic, cardiovascular and anthropometric traits (Metabolic/Cardio/Anthropometric), Neurological disorders, Psychiatric disorders.

(original cell atlas)). Concordantly, we observed that known PD-risk genes<sup>16</sup> are generally more highly expressed in neuronal cell types across the SN (Supplementary Fig. 7).

Our SN atlas associates the genetic risk of neuropsychiatric disorders with DaN gene expression but also with GABAergic gene expression, e.g. schizophrenia (SCZ) (DaN and GABA) (Fig. 2a). However, the conditional analyses between cell types of the SN demonstrated that where DaN and GABAergic neurons are both associated to a neuropsychiatric disorder, the association with DaNs is lost once conditioned upon GABA neuronal expression but not vice versa (Supplementary Table 6). Thus, the genetic risk of these psychiatric disorders is more broadly associated with genes expressed in the GABAergic neurons of the pars reticulata than the DaNs of the pars compacta. As previously reported<sup>17</sup>, we found in the cortex significant neuropsychiatric disorder cell-type associations with both Ex and In neurons (e.g. SCZ) (Supplementary Table 7) and that these associations are most often distinct from SN neurons (conditional

analysis; see Supplementary Table 8) proposing a distinct role of the SN in the aetiology of neuropsychiatric disorders, especially SCZ. Furthermore, we found a significant oligo-type association for different neuropsychiatric disorders (e.g. SCZ risk with two SN glial cell populations, namely OPC ( $q_{LDSC} = 4.27 \times 10^{-3}$ ,  $q_{MAGMA} = 1.32 \times 10^{-4}$ ) and ODC ( $q_{MAGMA} = 1.36 \times 10^{-5}$ )) (Fig. 2a). These associations support the hypothesis that for many brain disorders, glia may causally contribute to the neuronal alterations. This is well illustrated by the association of AD risk variants with microglia-expressed genes, found here for microglia in both the SN ( $q_{MAGMA} = 9.81 \times 10^{-4}$ ) and the cortex ( $q_{LDSC} = 0.01$ ,  $q_{MAGMA} = 7.2 \times 10^{-4}$ ) (Fig. 2, Supplementary Tables 5 and 7) and indistinguishable between microglia populations from the two brain regions (Supplementary Table 8,  $p = 0.84$ ).

There are significant overlaps in the genetic risk between different neuropsychiatric disorders<sup>18</sup>. We examined whether a common cell-type association between two neuropsychiatric

disorders was associated with a shared genetic risk between those disorders by re-performing the cell-association analysis using the genetic risk of one disorder after conditioning on the genetic risk effect of the other disorder. In general, these conditional analyses suggest that an overlapping component of risk between these disorders is associated with OPCs while distinct non-overlapping fractions of genetic risk act through the same neuronal types (Fig. 3). In some cases, we noted that a cell-specific association disappeared non-reciprocally when conditioning the risk on other disorders (e.g. GABAergic neurons and ADHD or Ex neurons and anxiety; Fig. 3), reflecting a more restricted and subsumed cell-specific genetic risk association for that disorder.

Finally, we also found genes specifically expressed in SN cell types to be associated with traits not specific to the brain. We identified a unique cell-type-specific signal between microglia and genetic variants associated with high-density lipoprotein (HDL) cholesterol level ( $q_{\text{MAGMA}} = 0.046$ ) (Fig. 2a). Given the role of lipids in AD<sup>19,20</sup>, we asked whether the HDL cholesterol association was related to the association between AD risk and microglia gene expression (Fig. 2a). However, conditional analysis suggests distinct genetics underlie the associations of AD and HDL cholesterol levels with microglia (Supplementary Fig. 8).

**Genetic risk highlights cell-type-specific networks and pathways.** Cell-type-specific gene expression patterns within the SN and cortex may help identify cellular circuitry that underlies disease-associated cell-type-specific vulnerabilities. We used protein-protein interactions (PPI) as evidence for functional relationships between genes, and built cell-type-specific PPI networks within which we identified a number of modules of highly interconnected genes (“Methods” and Supplementary Note 4). For each significant cell-type/disease-risk association discovered above (Fig. 2), we looked to refine the disease-risk association to a relevant cell-type-specific gene module using MAGMA gene set analysis (Fig. 4). PD risk showed association to nigral DaN modules (M1 and M2) (Fig. 4b, Supplementary Fig. 9 and Fig. 5) enacting processes previously associated with PD such as mitochondrial organisation and functioning, endocytosis, protein ubiquitination and macroautophagy<sup>21</sup>. PD-risk association to ODCs (M5) was related to genes enriched in metabolic processes, gene regulation, kinase activity, protein phosphorylation and neurogenesis, while the OPC risk-associated module (M1) was enriched in metabolic processes, gene regulation and cell differentiation. The SCZ-risk association to cortical neuronal populations was mainly related to synaptic signalling and neuronal developmental processes (Ex neuron module M1/In neuron M1 and Ex neuron module M4/In neuron M4, respectively; Fig. 4a and Supplementary Fig. 9). In the nigral neuronal population, genes within DaN module M4 and GABA module M3, both associated with synaptic signalling/neuron development, were also enriched in SCZ risk (Fig. 4b and Supplementary Fig. 9). A further SCZ association to DaN (M1) was found related to mitochondria (Fig. 4b), supporting the role of mitochondrial dysfunction in SCZ<sup>22</sup>. Lipid metabolism and neuron development were found enriched in both cortical ODCs (M2) and OPCs (M4) modules associated to SCZ risk. Furthermore, nigral OPC modules (M3 and M5), related to nervous system development and synaptic signalling, were found associated with SCZ risk (Fig. 4b and Supplementary Fig. 9). The nigral OPC M3 module also showed a significant association with other traits such as neuroticism (corrected  $p$  value = 0.004). As predicted by the genetic risk, a BP association was found to the nigral DaN module (M4) associated with synaptic signalling and neuronal development (Fig. 4b and Supplementary Fig. 9).

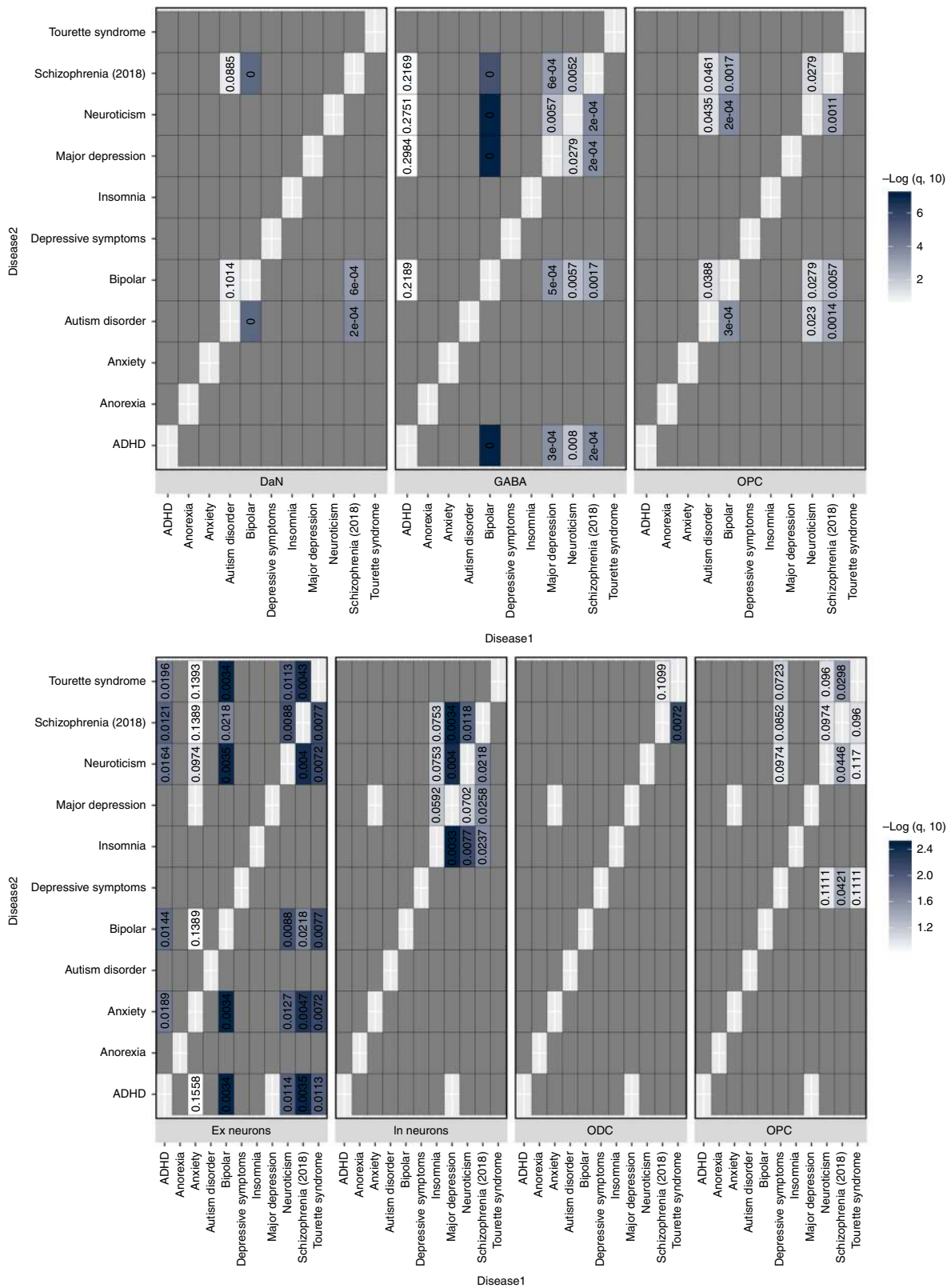
We then evaluated whether a common module association within the same cell-type between two disorders was associated with a shared genetic risk between those disorders. For this, we re-performed the cell-specific module-association analysis using the genetic risk of one disorder after conditioning on the genetic risk effect of the other disorder and vice versa. For the three nigral modules associated with two disorders (Fig. 4), we found the associations of PD and SCZ to DaN module M1 were independent ( $p$  value = 0.0002 for PD conditioned on SCZ and  $p$  value = 0.0009 for SCZ conditioned on PD) with PD-associated M1 genes mainly enriched in endocytosis processes (regulation of endocytosis, GO:0030100  $p$  value = 0.0003) and SCZ-associated M1 genes mainly enriched in neuronal migration and development (neuron migration, GO:0001764  $p$  value = 0.0004). Similarly, we found the associations between BP and SCZ to DaN M4 to be independent ( $p$  value = 0.0003 for BP conditioned on SCZ and  $p$  value = 0.002 for SCZ conditioned on BP) with BP-associated M4 genes mainly enriched in synaptic transmission processes (synaptic signalling, GO:0099536  $p$  value =  $3.9 \times 10^{-5}$ ) and SCZ-associated M4 genes mainly enriched in neuronal development (nervous system development, GO:0007399  $p$  value =  $6.3 \times 10^{-6}$ ). By contrast, we found that association of the nigral ODC M5 module with PD risk was lost when conditioned on SCZ risk ( $p$  value = 0.23) but that the SCZ association remained when conditioned on PD risk ( $p$  value = 0.002) revealing convergence, specifically that the SCZ-risk loci overlapped and subsumed the loci underlying the ODC M5 PD association.

## Discussion

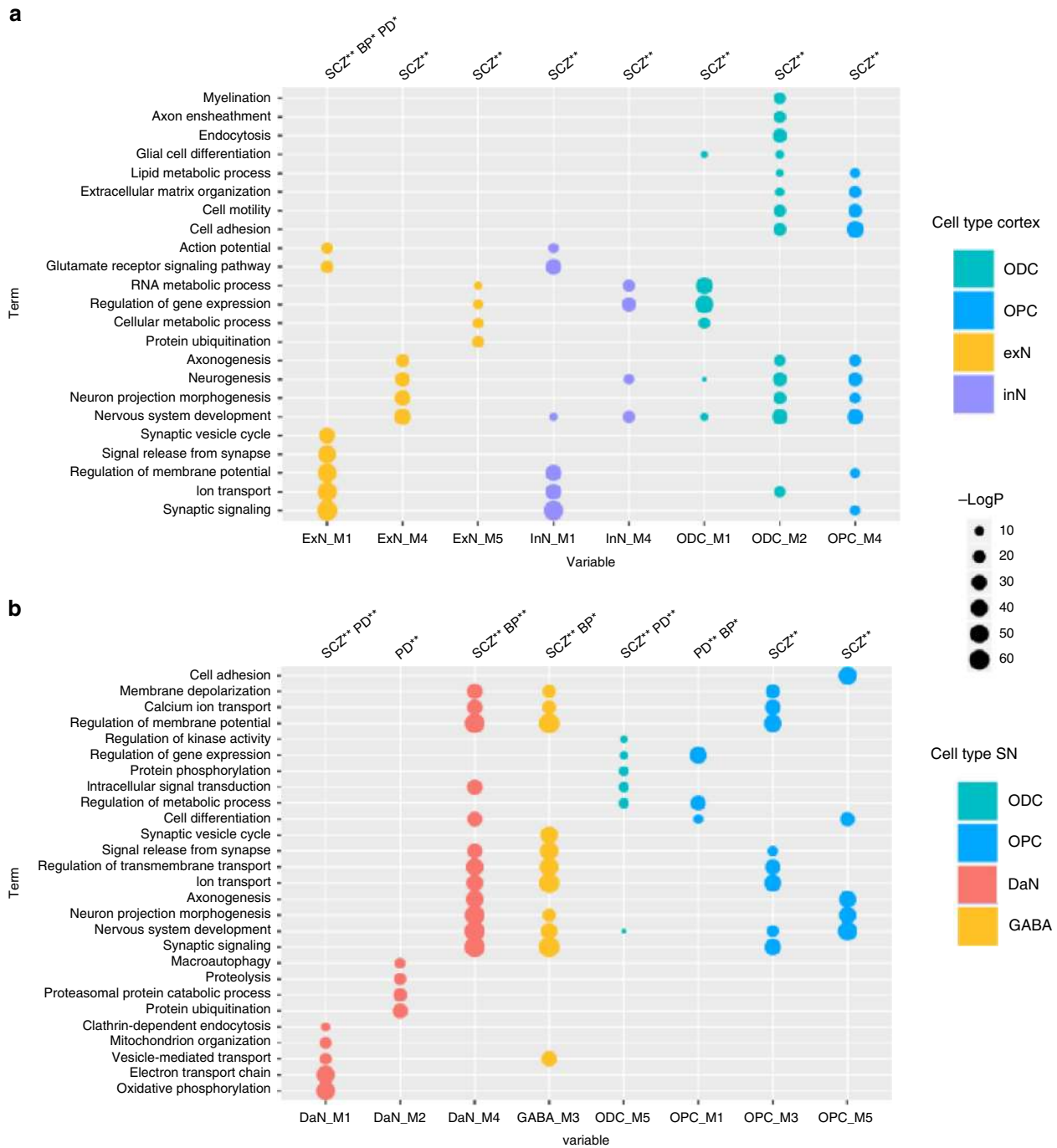
Here, we report a comprehensive human single-nuclei transcriptomic atlas for SN by sequencing ~17,000 nuclei from matched cortical and SN tissue enabling the identification and comparison of brain region-specific cell types. In this study, (i) we re-identify all the major cell types previously reported indicating sufficient genes/cells coverage, (ii) we find a large difference in the neuronal-glia cell composition between cortex and SN that confounds bulk tissue disease associations (Fig. 1, Supplementary Fig. 10 and Supplementary Table 9), (iii) we identify multiple associations between the genetic risk of particular diseases and specific cell types in the nigra and cortex (Fig. 2), (iv) we find that where there are multiple neuropsychiatric disease associations for a given neuron type, the loci associated for each disorder are distinct and do not appear to converge on the same set of genes within that cell type, while for glia we do observe convergence (Fig. 3) and (v) we determine the different cell-specific gene-networks and their functions perturbed by disease-risk variants (Figs. 4 and 5).

This SN/cortex paired-sample atlas is a valuable resource to interpret the genetic architecture of many disorders, especially for PD and other disorders with particular vulnerabilities in the SN as compared to the cortex. However, the fraction of PD genetic risk that mapped to cortical Ex neurons was indistinguishable from that mapping to SN DaNs, and thus our atlas alone is unable to propose why SN DaNs is lost earlier in PD pathology. Our cellular atlas suggests similar cellular functions associated with PD genetic risk across neuronal cell types in PD nigral and cortical degeneration<sup>2</sup>. Thus, different metabolic demands and different local environments of these neurons from different brain regions likely contribute to their distinct vulnerabilities<sup>23</sup>. PD genetic risk also appears to manifest through ODCs and OPCs (Fig. 2), which implicate gene regulation in metabolic processes and in cell development and support the growing evidence of the role of glia in neurodegenerative disorders<sup>24</sup>. Notably, we observe expression of the known PD gene *LRRK2* to be significantly higher in OPCs than other SN cell types (Supplementary Fig. 7).





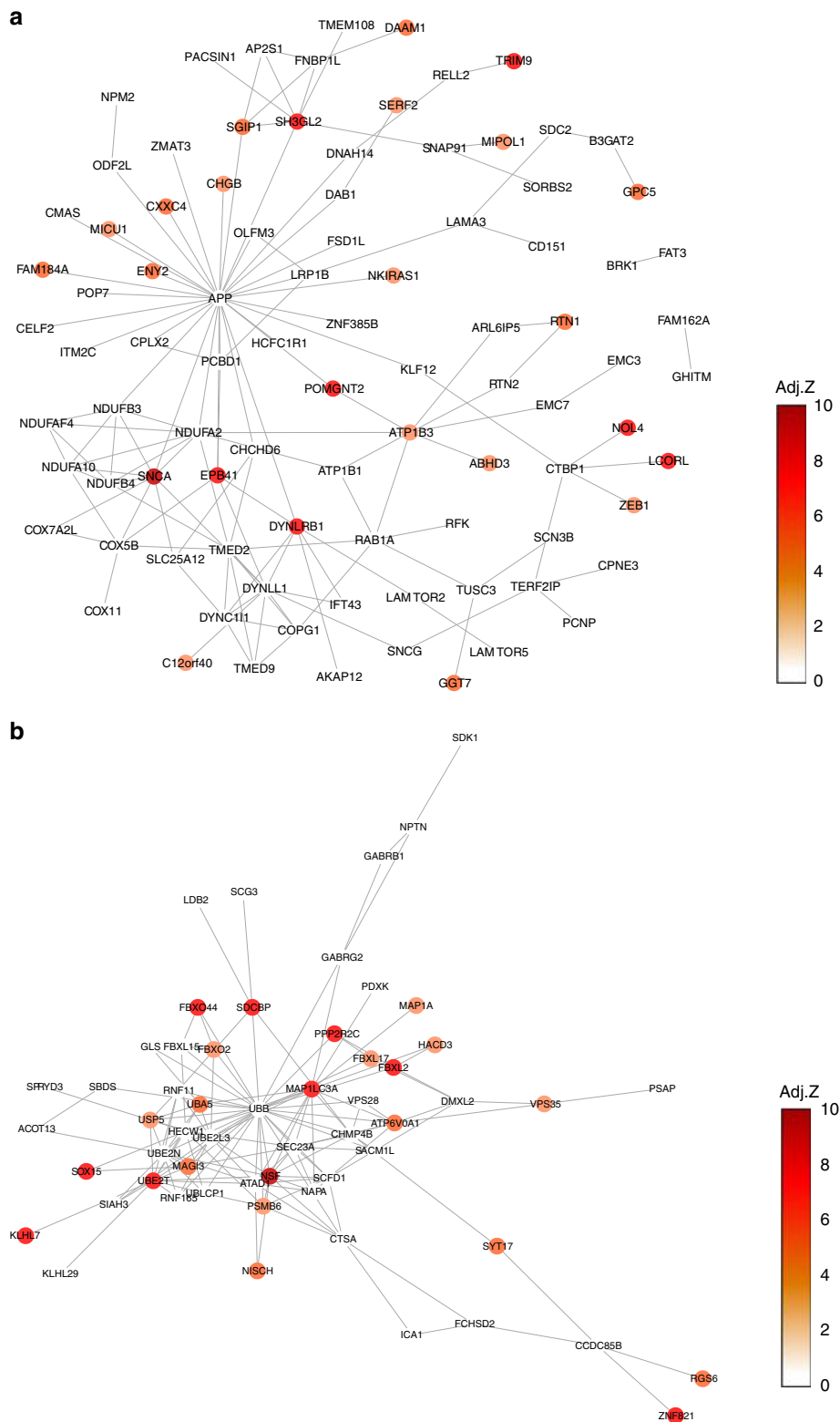
**Fig. 3 Evaluation of the shared cell-type associations between pairs of neuropsychiatric disorders.** Evaluation of alike cell-type associations between any two neuropsychiatric disorders to identify any shared cell-type-specific component of risk, for the SN cell types (top) and the cortex (bottom). Each heatmap represents the results from LDSC of the associations ( $p$  value associated with an LDSC coefficient) of a specific cell-type expression profile with the genetic risk of a given neuropsychiatric disorder (disease1—X-axis) after conditioning on the genetic risk of another neuropsychiatric disorder (disease2—Y-axis). This analysis was only performed where two neuropsychiatric disorders showed a significant (or suggestive) association with the same cell type (Fig. 2). The blue heatmap colours are proportional to  $-\log_{10} q$  value (FDR-adjusted  $p$  value) of the enrichment of genetic variants associated with a disorder adjusted for another disorder. The cell associations that were not evaluated (no overlap in Fig. 2) are coloured in dark grey.



**Fig. 4 Cell-type-specific pathways associated with the genetic risk of three genetic brain disorders.** Gene Ontology (GO) enrichment for cortex (a) and substantia nigra (SN) (b) cell-type-specific protein-protein interaction (PPI) gene modules enriched in Parkinson’s disease (PD), schizophrenia (SCZ) and bipolar (BP) disorders risk. We tested the convergence of disease genetic risk at a functional level across cell-type-specific PPI gene modules using MAGMA gene set analysis (one-sided positive two-sample *t*-test); an asterisk (\*) and double asterisks (\*\*) indicate nominally significant *p* value (<0.05) and *q* value (Bonferroni correction for the number of cell-type PPI modules tested), respectively. The top representative GO biological process terms are shown for cell-type modules with either PD, SCZ or BP risk enrichment that reached significance in the more general cell-type level analysis (Fig. 2). Tests across all identified gene modules are reported in Supplementary Fig. 10, and complete lists of enriched GO terms are reported in Supplementary Data 6 and 7. The size of circles represents  $-\log(p \text{ value})$  for GO enrichment with Fisher test; colours correspond to cell types (DaNs dopaminergic neurons, Ex excitatory neurons, GABA GABAergic neurons, In: inhibitory neurons, ODC oligodendrocytes, OPC oligo-precursor cells).

A role for ODCs in PD is surprising given the light myelination of SN DaN axons<sup>25</sup>, but our relatively unbiased approach allows for unexpected associations that require further understanding. Indeed, multiple studies have identified white matter impairments, which correlate with progression and appear to precede

grey matter atrophy, and our results may provide a missing molecular link to PD genetic risk<sup>26</sup>. The association we discover between PD genetic risk variants and human ODC-specific gene expression has very recently been proposed in an independent study<sup>13</sup> using mouse transcriptomic data. Unlike AD, we do not



**Fig. 5 Cell-type-specific protein-protein interaction networks of Parkinson's risk-associated genes.** Protein-protein interaction (PPI) network representations of Parkinson's risk-associated genes within substantia nigra (SN) dopaminergic neurons (DaN) specific modules DaN\_M1 (**a**) and DaN\_M2 (**b**). We display all the genes found within each of two DaN-specific PPI gene modules associated with Parkinson's disease risk (Fig. 4) and colour each gene according to that gene's Parkinson's risk association MAGMA adjusted Z-score, with darker colours corresponding to higher adjusted Z-scores and a stronger Parkinson's risk association. Edges between genes indicate an interaction between those genes' protein products (see "Methods").



find an association between PD genetic risk and microglia, suggesting neuroinflammation may play a lesser role in PD risk than in AD risk.

The midbrain encompassing the SN is a key brain region of interest after the cortex for several neuropsychiatric disorders<sup>27</sup>. For SCZ we found distinct associations with both nigral neurons and cortical neurons. However, it may be that this SCZ-DaN association relates to DaNs within the adjacent ventral tegmental area (VTA) region previously implicated in several neuropsychiatric disorders<sup>28</sup>. Without corresponding VTA atlases from these brains, we are unable to perform the relevant conditional tests, which highlights the value of larger future studies capturing more regions from the same brains. Due to the key role of DaNs in PD, the role of GABAergic neurons of the pars reticulata in other diseases is relatively understudied. Nonetheless, the GABAergic neurons project to the prefrontal cortex and nucleus through mesolimbic pathway and control mainly the reward system (learning about motivationally relevant stimuli in the environment), which is affected in patients with neuropsychiatric disorders<sup>29</sup> and our novel associations suggest these neurons have the potential to be influenced by the genetic risk of several neuropsychiatric disorders. Across several neuropsychiatric and neurodegenerative disorders, we found multiple glia association, especially to OPC processes related to synaptic signalling<sup>30</sup>.

We describe here the first comprehensive human SN cell-type atlas. Together with a matching cortical atlas, these atlases allow the systematic characterisation of SN cell types and the identification of cell-type-specific processes influenced by multiple complex disorders.

## Methods

**Samples.** Five controls were selected on the basis of the absence of neurological clinical disease and by midbrain RNA integrity number yielding scores over eight from the Oxford Brain Bank (Supplementary Table 1). However, the histopathological examination revealed cerebral amyloid angiopathy (CAA) in the brain of one individual (Sample 3, Supplementary Table 1). CAA is one of the morphologic hallmarks of AD<sup>31</sup>, but it is also, very common in the brains of elderly patients who are neurologically healthy<sup>32</sup>. To ensure that this individual did not affect the results and conclusions of this study, we repeated all further analyses with and without the samples from this individual and found a very high correlation between results with highly concordant cell clustering and thus retained this sample for greater power. Informed consent had been collected from all cases fulfilling the requirements of the Human Tissue Act 2004. Blocks of ~3 mm<sup>2</sup> were dissected from snap-frozen slices by a qualified neuropathologist (J.A.-A.) from the central portion of the SN at the level of the third nerve encompassing both ventral and dorsal tiers and from the middle frontal gyrus encompassing the cortex but macroscopically excluding the subcortical white matter. Replicate blocks obtained on a different day were obtained in two out of the five individuals.

**10x sequencing data.** 10x chromium single-nuclei sequencing was performed on the cortex and SN regions from five individuals. In total 12 samples (including 2 SN replicates), with a total of 12,015 and 6105 nuclei, were sequenced for the cortex and SN over 2 days (Supplementary Table 1). Reads were processed and mapped to the Human Genome (GRCh38.84-premrna) with Cell Ranger 2.1.1 (Supplementary Note 5). We recovered a median number of 2455 and 690 nuclei after sequencing for the cortex and SN. The multiple sample libraries were sequenced with mean reads ranging from 46,598 to 59,513 and 18,377 to 44,710 for the cortex and SN, respectively. Overall we detected median genes per nuclei ranging from 607 to 3364 for the multiple individuals across both regions (Supplementary Data 1). Similar in range as other human brain single-nuclei studies<sup>6,33</sup>, the mean read depth per nuclei was 43,150 (95% CI: 34,488–51,821) and the median number of genes detected per nucleus was 1886 (95% CI: 1228–2544) (Supplementary Data 1). For the SN single-nuclei atlas, we generated biological replicates and showed that cell profiles from the same individual were comparable to each other (Supplementary Fig. 11).

**Data processing.** The filtered unique molecular identifiers (UMI) feature-barcode matrices generated with CellRanger 2.1.1, were processed with Seurat R package (v2.3.4)<sup>34</sup>, separately for the SN and the Cortex. As quality control steps, we retained genes with a count of 1 in at least three nuclei and removed nuclei with <500 genes per sample, high thresholds of nUMI (range of >5000–12000, sample

dependent), mitochondrial percentage >0.05 and ribosomal percentage >0.05. The distributions of number of genes, number of UMIs and percentage of reads mapped to mitochondrial and ribosomal genome were further inspected for quality assurance. Each normalised sample was linearly regressed to remove any inter-cellular gene expression variation because of technical effects associated with UMI coverage, ribosomal and mitochondrial percentage, followed by gene level scaling of the data by using the “ScaleData” function in Seurat. Cell cycle phase scores were predicted for each cell per sample and determined to not be an important source of variation and bias in the SN and cortex (Supplementary Figs. 1 and 3).

**Cell clustering analysis.** After merging the samples, identifying common sources of variation based on the highly variable genes (Supplementary Note 6), performing a canonical correlation analysis (CCA)<sup>34</sup>, and discarding rare non-overlapping cells between samples, 5943 nuclei and 22,736 genes (15,568 protein-coding genes) remained for the SN and 10,706 nuclei and 26,145 genes (16,423 protein-coding genes) remained for the cortex. The CCA analysis identified the top numbers of CCA vectors to align for the SN and cortex as 25 and 42 dimensions. The shared-nearest neighbour graph was constructed on a cell-to-cell distance matrix from the top aligned CCA vectors followed by Louvain clustering<sup>35</sup> to identify cell-type clusters, which were visualised with t-distributed stochastic neighbour embedding and uniform manifold approximation and projection plots.

On the basis of previous knowledge, consistency and validity of the different resolutions, we selected the final number of clusters based on the clustering resolution for the Louvain algorithm in the range of 0.4–0.8, which included all the major cell types and subtypes (Supplementary Notes 6 and 7). Cell clusters with fewer than 30 cells were omitted from further analysis.

**Cell-type annotation for cortex and SN.** We characterised the cellular identities of clusters in the SN (Supplementary Notes 1 and 7, Fig. 1a, c) and the cortex (Supplementary Notes 1 and 7, Fig. 1b) nuclei by identifying known marker genes enriched in each of the clusters (Supplementary Notes 1 and 7)<sup>6,36</sup>. For each of the subclusters, the enriched marker genes were identified by differential expression of the cells grouped in each sub-cluster against the remaining cells within the corresponding broad cell-type cluster. This process resulted in the annotation of 23 and 10 different neuronal and non-neuronal cell types for the cortex and SN, respectively (Fig. 1a, b).

**Differential expression.** Differentially expressed genes between cell types and subtypes were identified within Seurat<sup>34</sup> by using the negative binomial test (false discovery rate (FDR)-corrected  $p$  value <0.05) to identify 0.25 log fold enriched genes detected in at least 25% of cells in the cluster of interest. Differential expression analyses were performed for all clusters separately in the SN and cortex, for SN astrocyte and DaN subtypes and for cortical Ex and In neuronal subtypes (Supplementary Data 3).

**Comparison of 10x data sn-RNAseq data with published data.** Comparison of the cortex and SN nuclei averaged cell-type populations with previously published single-cell human temporal cortex<sup>6</sup>, single nuclei<sup>5</sup> and bulk SN laser-capture microscopy (LCM) DaNs<sup>7</sup> and glial cell types<sup>8</sup> data further confirmed the broad cell-type annotations and, in the case of cortex, neuronal subtype classification consistency (Supplementary Fig. 12, Supplementary Note 8). In particular, SN DaNs identified in our 10x study cluster with external LCM postmortem DaNs (Supplementary Fig. 12) and astrocyte-1 or reactive astrocyte population cluster with the external LCM astrocyte and microglia samples (Supplementary Fig. 12). In addition, cell types are comparable between single-nuclei and single-cell data (Supplementary Fig. 12). A correlation >0.85 was observed between all the major cortical cell types and layer-specific neuronal subtypes between single-nuclei cell atlas interrogated in this study and the human visual cortex and frontal cortex single-nuclei-drop seq data<sup>5</sup> (Supplementary Fig. 12). The data processing and method of comparison for each of the external datasets are described in Supplementary Information.

**Cell-type-association analysis.** We intersected cell-type-specific expression patterns with genetic risk of specified disease to identify disease-relevant cell types in SN and cortex for 30 diseases and traits (Fig. 2, Supplementary Table 4, Supplementary Note 4). We performed these cell-type association analyses with two commonly used approaches: LDSC<sup>11</sup> and MAGMA<sup>12</sup>. While these methods are complementary, we confirmed that their findings were well aligned (Supplementary Fig. 13, Supplementary Note 4). Where we identified multiple cell types associated with the same trait, we performed conditional analyses to evaluate whether it was the same set of genetic variants acting in different cell types or distinct sets of genetic variants in each cell-type suggesting multiple cellular aetiologies. Where traits/disorders were associated with both SN and cortical cell types, we considered that the different proportions of glial versus neurons for these two brain regions could bias the identification of cell-type-specific genes. To test this potential bias, we repeated the conditional analysis with artificially homogeneous SN/cortex cellular atlas, which included the same proportion of glial cells (Supplementary Note 4). To evaluate the genetic overlap between two traits, each showing an association with the same cell type, we performed GWA conditional analysis with

multi-trait-based conditional and joint analysis<sup>37</sup> (Fig. 3). We then repeated LDSC analyses with the GWA summary statistic of one trait adjusted for the second trait and vice versa.

**Functional analysis.** *Cell-type-specific PPI network and identification of gene modules:* A cell-type-specific PPI network is built by extracting PPIs from PPI network between cell-type-specific genes (Supplementary Data 4 and Supplementary Note 4). To identify modules of highly interconnected genes in a cell-type-specific PPI network, we employed “cluster\_louvain” function in “igraph” R package<sup>38</sup>. This function implements the multi-level modularity optimisation algorithm, where at each step genes are re-assigned to modules in a greedy way and the process stops when the modularity does not increase in a successive step. Modules with >30 genes are used for further analysis as smaller modules have low informativity. MAGMA gene set analysis was used to test enrichment in disease-risks across all identified modules.

*Gene Ontology (GO) pathway enrichment analysis:* We performed GO enrichment analysis with topGO R package by testing the overrepresentation of GO biological processes (GO BP) terms within the input gene sets using Fisher test. Revigo<sup>39</sup> was used to summarise the top 100 enriched GO BP terms in a smaller number of categories.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

### Data availability

The processed 10x 3' Chromium single-nuclei RNAseq UMI-barcode matrices for each sample are available from the Gene Expression Omnibus under the accession code GSE140231.

Received: 11 December 2019; Accepted: 21 July 2020;

Published online: 21 August 2020

### References

- Hickman, S., Izzy, S., Sen, P., Morsett, L. & El Khoury, J. Microglia in neurodegeneration. *Nat. Neurosci.* **21**, 1359–1369 (2018).
- Poewe, W. et al. Parkinson disease. *Nat. Rev. Dis. Prim.* **3**, 17013 (2017).
- Booth, H. D. E., Hirst, W. D. & Wade-Martins, R. The role of astrocyte dysfunction in Parkinson's disease pathogenesis. *Trends Neurosci.* **40**, 358–370 (2017).
- Bruck, D., Wenning, G. K., Stefanova, N. & Fellner, L. Glia and alpha-synuclein in neurodegeneration: a complex interaction. *Neurobiol. Dis.* **85**, 262–274 (2016).
- Darmanis, S. et al. A survey of human brain transcriptome diversity at the single cell level. *Proc. Natl Acad. Sci. USA* **112**, 7285–7290 (2015).
- Lake, B. B. et al. Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. *Nat. Biotechnol.* **36**, 70–80 (2018).
- Skene, N. G. et al. Genetic identification of brain cell types underlying schizophrenia. *Nat. Genet.* **50**, 825–833 (2018).
- Liddel, S. A. & Barres, B. A. Reactive astrocytes: production, function, and therapeutic potential. *Immunity* **46**, 957–967 (2017).
- von Bartheld, C. S., Bahney, J. & Herculano-Houzel, S. The search for true numbers of neurons and glial cells in the human brain: a review of 150 years of cell counting. *J. Comp. Neurol.* **524**, 3865–3895 (2016).
- Donovan, M. K. R., D'Antonio-Chronowska, A., D'Antonio, M. & Frazer, K. A. Cellular deconvolution of GTEx tissues powers discovery of disease and cell-type associated regulatory variants. *Nat. Commun.* **11**, 955 (2020).
- Finucane, H. K. et al. Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat. Genet.* **50**, 621–629 (2018).
- de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.* **11**, e1004219 (2015).
- Bryois, J. et al. Genetic identification of cell types underlying brain complex traits yields insights into the etiology of Parkinson's disease. *Nat. Genet.* **52**, 482–493 (2020).
- Nalls, M. A. et al. Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet Neurol.* **18**, 1091–1102 (2019).
- Reynolds, R. H. et al. Moving beyond neurons: the role of cell type-specific gene regulation in Parkinson's disease heritability. *NPJ Parkinsons Dis.* **5**, 6 (2019).
- Billingsley, K. J., Bandres-Ciga, S., Saez-Atienzar, S. & Singleton, A. B. Genetic risk factors in Parkinson's disease. *Cell Tissue Res.* **373**, 9–20 (2018).
- Lake, B. B. et al. Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. *Science* **352**, 1586–1590 (2016).
- Brainstorm, C., et al. Analysis of shared heritability in common disorders of the brain. *Science* **360**, eaap8757 (2018).
- Button, E. B. et al. HDL from an Alzheimer's disease perspective. *Curr. Opin. Lipido.* **30**, 224–234 (2019).
- Liao, F., Yoon, H. & Kim, J. Apolipoprotein E metabolism and functions in brain and its role in Alzheimer's disease. *Curr. Opin. Lipido.* **28**, 60–67 (2017).
- Cherubini, M. & Wade-Martins, R. Convergent pathways in Parkinson's disease. *Cell Tissue Res.* **373**, 79–90 (2018).
- Srivastava, R., Faust, T., Ramos, A., Ishizuka, K. & Sawa, A. Dynamic changes of the mitochondria in psychiatric illnesses: new mechanistic insights from human neuronal models. *Biol. Psychiatry* **83**, 751–760 (2018).
- Fu, H., Hardy, J. & Duff, K. E. Selective vulnerability in neurodegenerative diseases. *Nat. Neurosci.* **21**, 1350–1358 (2018).
- Ettle, B., Schlachetzki, J. C. M. & Winkler, J. Oligodendroglia and myelin in neurodegenerative diseases: more than just bystanders? *Mol. Neurobiol.* **53**, 3046–3062 (2016).
- Haddad, D. & Nakamura, K. Understanding the susceptibility of dopamine neurons to mitochondrial stressors in Parkinson's disease. *FEBS Lett.* **589**, 3702–3713 (2015).
- Duncan, G. W. et al. Gray and white matter imaging: a biomarker for cognitive impairment in early Parkinson's disease? *Mov. Disord.* **31**, 103–110 (2016).
- Hannon, E., Marzi, S. J., Schalkwyk, L. S. & Mill, J. Genetic risk variants for brain disorders are enriched in cortical H3K27ac domains. *Mol. Brain* **12**, 7 (2019).
- Arrondo, G. et al. Reduction in ventral striatal activity when anticipating a reward in depression and schizophrenia: a replicated cross-diagnostic finding. *Front. Psychol.* **6**, 1280 (2015).
- Gold, J. M., Waltz, J. A., Prentice, K. J., Morris, S. E. & Heerey, E. A. Reward processing in schizophrenia: a deficit in the representation of value. *Schizophr. Bull.* **34**, 835–847 (2008).
- Sakry, D. et al. Oligodendrocyte precursor cells modulate the neuronal network by activity-dependent ectodomain cleavage of glial NG2. *PLoS Biol.* **12**, e1001993 (2014).
- Yamaguchi, H., Yamazaki, T., Lemere, C. A., Frosch, M. P. & Selkoe, D. J. Beta amyloid is focally deposited within the outer basement membrane in the amyloid angiopathy of Alzheimer's disease. An immunoelectron microscopic study. *Am. J. Pathol.* **141**, 249–259 (1992).
- Yamada, M. Risk factors for cerebral amyloid angiopathy in the elderly. *Ann. N. Y. Acad. Sci.* **977**, 37–44 (2002).
- Mathys, H. et al. Single-cell transcriptomic analysis of Alzheimer's disease. *Nature* **570**, 332–337 (2019).
- Butler, A., Hoffman, P., Smibert, P., Papalexis, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).
- Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech.* **2008**, P10008 (2008).
- Mancarci, B. O. et al. Cross-laboratory analysis of brain cell type transcriptomes with applications to interpretation of bulk tissue data. *eNeuro* **4**, ENEURO.0212-17.2017 (2017).
- Zhu, Z. et al. Causal associations between risk factors and common diseases inferred from GWAS summary data. *Nat. Commun.* **9**, 224 (2018).
- Csardi, G. & Nepusz, T. The igraph software package for complex network research. *InterJournal, Complex Syst.* **1695**, 1–9 (2006).
- Supek, F., Bosnjak, M., Skunca, N. & Smuc, T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS ONE* **6**, e21800 (2011).

### Acknowledgements

C.W., V.V. and C.S. are supported by the UK Dementia Research Institute (UK DRI) funded by the Medical Research Council (MRC), Alzheimer's Society and Alzheimer's Research UK (AR-UK). C.S. is supported by the Ser Cymru II programme, which is part-funded by Cardiff University and the European Regional Development Fund through the Welsh Government. D.A. was supported by the AR-UK Oxford Drug Discovery Institute, University of Oxford. We acknowledge the Oxford Brain Bank, supported by the Medical Research Council (MRC), Brains for Dementia Research (BDR) (Alzheimer Society and Alzheimer Research UK), and the NIHR Oxford Biomedical Research Centre. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. This work was supported by the Wellcome Trust via core funding to the Wellcome Centre for Human Genetics (award 203141/Z/16/Z). R.W.-M., J.A.-A. and T.M.C. were supported by the Monument Trust Discovery Award from Parkinson's UK. R.B. was supported by the Wellcome Trust Centre for Human Genetics, University of Oxford, UK.

### Author contributions

D.A., C.S. and V.V. performed data analyses and wrote paper, T.M.C. developed and performed nuclei extraction protocols, J.M.-S. performed data analyses, RB performed sequencing and generated data and contributed to experimental design, J.A.-A. contributed to experimental design and the selection and sourcing of postmortem tissue, R.W.-M. developed nuclei extraction protocols and contributed to experimental design and C.W. conceived project, designed experiments, designed and led analysed and wrote paper.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41467-020-17876-0>.

**Correspondence** and requests for materials should be addressed to C.W.

**Peer review information** *Nature Communications* thanks Tristan Qingyun Li and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020