

1 **TITTLE PAGE**

2 **LCM-seq identifies robust markers of vulnerable and resistant human midbrain dopamine neurons**

3

4 Julio Aguila^{1,#}, Shangli Cheng^{2,3,#}, Nigel Kee¹, Ming Cao¹, Qiaolin Deng^{2,3,*} and Eva Hedlund^{1,4,*}

5 ¹ Department of Neuroscience, Karolinska Institutet, Stockholm, 171 77, Sweden

6 ² Department of Physiology and Pharmacology, Karolinska Institutet, Stockholm, 171 77, Sweden

7 ³ Center for Molecular Medicine, Karolinska University Hospital, Stockholm, 171 77, Sweden

8 ⁴ Lead contact

9 #Equal contribution

10 *Correspondence: eva.hedlund@ki.se (E.H.) or qiaolin.deng@ki.se (Q.D.)

11 **Running title: Novel markers of midbrain dopamine neuron groups**

12

13 **Keywords**

14 Human midbrain dopamine neurons, spatial transcriptomics, laser microdissection, RNA sequencing,
15 substantia nigra compacta, ventral tegmental area, Parkinson Disease

16

17

18

19

20

21

22

23

24

25

26

1 **ABSTRACT**

2 Defining transcriptional profiles of substantia nigra pars compacta (SNc) and ventral tegmental area (VTA)
3 dopamine neurons in human is critical to understanding their differential vulnerability in Parkinson Disease.
4 However, reported marker profiles for these neuron populations are derived predominantly from rodents,
5 utilize small sample sizes and display extensive variability between studies. Here, we map selective
6 expression profiles of dopamine neurons in an extensive collection of human SNc and VTA using laser
7 capture microdissection coupled with Smart-seq2 RNA sequencing (LCM-seq). By applying a
8 bootstrapping strategy as sample input to DESeq2, we identify 33 differentially expressed SNc- or VTA-
9 specific markers and we also compute the minimal cohort size required to identify differentially expressed
10 genes (DEGs) that are concordant regardless of cohort size. Among the identified DEGs, *ZCCHC12*,
11 *CDH13* and *SERPINE2*, are minimally required to distinguish SNc or VTA dopamine neurons in both
12 human and mouse. In summary, our study identifies novel markers, besides previously identified ones,
13 which will be instrumental for future studies aiming to modulate dopamine neuron resilience as well as
14 validate cell identity of stem cell-derived dopamine neurons.

15

16 **INTRODUCTION**

17 Midbrain dopamine neurons are divided into two major populations, the substantia nigra pars compacta
18 (SNc) and the ventral tegmental area (VTA) (Hedlund and Perlmann 2009). SNc dopamine neurons project
19 to the dorsolateral striatum (Dahlstroem and Fuxe 1964) and are severely affected in Parkinson Disease
20 (PD) (Damier et al. 1999b; Damier et al. 1999a), while VTA dopamine neurons project to cortical and
21 mesolimbic areas and are more resilient to degeneration (Hedlund and Perlmann 2009). These neuron
22 populations have been extensively investigated in numerous rodent models, (Grimm et al. 2004; Chung et
23 al. 2005; Greene et al. 2005; Bifsha et al. 2014; Poulin et al. 2014), towards the goal of identifying molecular
24 mechanisms that can prevent degeneration or to model disease. Targeted analysis of midbrain dopamine
25 neuron populations have revealed several markers that appear to differentially label SNc e.g. *Aldh1a7*, *Sox6*,
26 *Cbln1*, *Vav3*, *Atp2a3* and VTA e.g. *Calb1*, *Otx2*, *Crym*, *Cadm1* and *Mareks* (Damier et al. 1999a; Grimm

1 et al. 2004; Chung et al. 2005; Greene et al. 2005; Di Salvio et al. 2010; Bifsha et al. 2014; Panman et al.
2 2014; Nichterwitz et al. 2016). Transcriptional analysis of human tissue has largely been limited to SNc
3 (Cantuti-Castelvetri et al. 2007; Simunovic et al. 2009) except for a recent small sample cohort of both SNc
4 and VTA (Nichterwitz et al. 2016). Frustratingly, these aforementioned investigations display extensive
5 cross-study variability, resulting in very few reproducible markers both within the same species and across
6 mouse, rat and human. Small sample sizes could be confounding these findings, along with differences in
7 rodent strain backgrounds, inter-individual variability among human patients, and methodological
8 differences. Thus, these discrepancies warrant further comprehensive cross-species comparative analyses.

9
10 To identify human SNc and VTA specific markers, which could reveal cell intrinsic properties underlying
11 their differential vulnerability in PD, a thorough large-scale transcriptional profiling in adult human tissues
12 is required. Such an analysis could also investigate the minimum cohort size necessary, above which lineage
13 specific markers remain stably differentially expressed irrespective of patient selection, an essential
14 requirement for valid study design in variable human populations. Finally, identified differences could also
15 serve as a foundation for the selective *in vitro* derivation of SNc dopamine neurons, which represent the
16 ideal cell type for transplantation in PD (Schultzberg et al. 1984; Haque et al. 1997; Thompson et al. 2005;
17 Hedlund and Perlmann 2009; Kriks et al. 2011; Ganat et al. 2012).

18
19 Here we used the spatial transcriptomics method LCM-seq, which combines laser capture microdissection
20 with Smart-seq2 RNA sequencing (Nichterwitz et al. 2016; Nichterwitz et al. 2018), to precisely analyze
21 individually isolated SNc and VTA dopamine neurons from 18 human *post mortem* brains. Using a
22 bootstrapping without replacement coupled with DESeq2, we identify 33 markers that were stably
23 differentially expressed between SNc and VTA dopamine neurons. We show that the minimal cohort size
24 required to reliably identify these subtype-specific markers were eight subjects, which may explain why
25 smaller cohorts have given inconsistent results. We further applied this approach to a large-scale data set
26 of mouse single midbrain dopamine neurons, revealing 89 SNc and VTA mouse stable genes. Finally, we

1 show that only three of the human stable genes, *ZCCHC12*, *CDH13* and *SERPINE2*, overlap with the mouse
2 stable genes, but that these three markers nevertheless faithfully classify SNc or VTA dopamine neurons in
3 both species. Several of the markers identified here have been implicated in PD or other degenerative
4 diseases and thus provide compelling future targets to modulate neuronal vulnerability or to model disease.

5

6 **RESULTS**

7 **Published inter- and intra-species SNc and VTA transcriptional profiles display considerable** 8 **discrepancies**

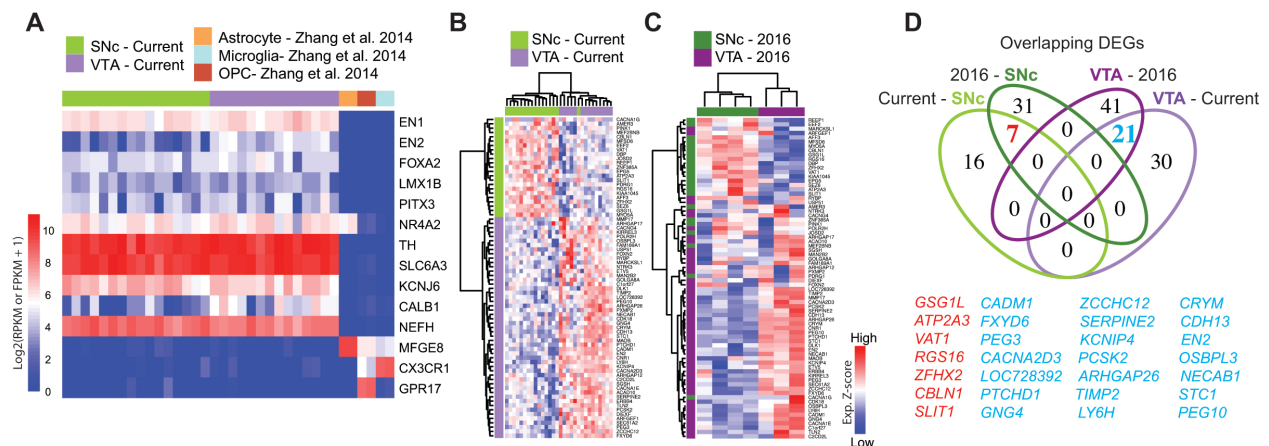
9 To begin constructing a consensus for VTA and SNc specific molecular profiles, we compiled previously
10 published transcriptome data from mouse and rat, for VTA or SNc differentially expressed genes (DEGs)
11 (Grimm et al. 2004; Chung et al. 2005; Greene et al. 2005; Bifsha et al. 2014; La Manno et al. 2016). This
12 analysis revealed that a surprisingly low fraction of DEGs were common across data sets (Supplemental
13 Fig. S1A, B; Supplemental Tables S1 and S4). Comparing across species with our previously published
14 small data set on human SNc and VTA (Nichterwitz et al. 2016), only two genes, *SOX6* and *CALBI*,
15 overlapped within SNc and VTA gene lists, respectively (Supplemental Fig. S1C). These discrepancies
16 highlighted the urgent need to identify reproducible marker profiles for VTA and SNc dopamine neurons.

17

18 **LCM-seq derives SNc or VTA specific markers from a large human cohort**

19 To identify robust and specific human dopamine neuron subpopulation markers, we isolated individual
20 VTA and SNc neurons from *post mortem* tissues of 18 adult individuals by LCM (Supplemental Fig. S2A-
21 G; Supplemental Table S2) and conducted polyA-based RNA sequencing. This study represents the largest
22 human data set profiling of SNc and VTA dopamine neurons to date. The quality of human fresh frozen
23 tissues used may vary as a consequence of *post mortem* interval (PMI), sample handling and preservation.
24 Therefore, prior to conducting differential gene expression analysis we performed extensive quality control
25 analysis to rule out undesired influences from sample processing (Supplemental Table S5). Randomly
26 selected samples that exhibited different PMIs for VTA or SNc neurons displayed comparable cDNA

1 quality (Supplemental Fig. S2H, I). Furthermore, while the total number of reads varied between individual
 2 samples, such variability was similarly distributed between SNc and VTA samples (Supplemental Fig. S2J).
 3 The number of detected genes did not correlate with either the age of the donor (Supplemental Fig. S2L),
 4 the PMI (Supplemental Fig. S2M) or the total number of reads (Supplemental Fig. S2K). Only the number
 5 of collected cells per sample modestly impacted gene detection ($P=0.515$) (Supplemental Fig. S2N).
 6 However, neither the number of collected cells nor the number of detected genes were significantly different
 7 between SNc and VTA neuron groups (Supplemental Fig. S2O, P) and thus should not affect DEG
 8 identification. Finally, we observed that all samples strongly expressed the dopamine neuron markers
 9 *EN1/2*, *FOXA2*, *LMX1B*, *PITX3*, *NR4A2*, *TH* and *SLC6A3* (DAT), and the general neuronal marker *NEFH*,
 10 while lacking glial markers *MFGE8*, *CX3CR1* or *GPR17*. This clearly demonstrates the selective
 11 enrichment of dopamine neurons using the LCM-seq methodology (Fig. 1A). *KCNJ6* (*GIRK2*) and *CALB1*,
 12 two genes often used to distinguish between SNc or VTA dopamine neurons, were also expressed (Fig.
 13 1A), but interestingly could not, on their own, accurately classify our samples (Supplemental Fig. S2Q).
 14



15 **Figure 1**

16 **Figure 1.** Gene signature of human adult midbrain dopamine neurons using LCM-seq. (A) The high sample
 17 quality for the 18 male subjects profiled in this study was confirmed by strong expression of the midbrain
 18 dopamine neuron markers *EN1/2*, *FOXA2*, *LMX1B*, *PITX3*, *NR4A2*, *TH* and *SLC6A3* (DAT), the pan-
 19 neuronal marker neurofilament (*NEFH*), and the lack of astrocyte, microglia or oligodendrocyte precursor

1 contamination (Zhang et al. 2014). (B) Hierarchical clustering analysis of samples from the current study
2 using the 74 DEGs identified by DESeq2. (C) The 74 DEGs also separated SNc and VTA samples from
3 Nichterwitz et al., 2016 (3 female subjects). (D) Venn-diagram displaying an unexpected low degree of
4 overlap between SNc and VTA DEGs found independently in each study. See also Supplemental Fig. S2.
5
6 Differential expression analysis, considering these 18 individuals, identified 74 DEGs (Supplemental Table
7 S6), which resolved SNc from VTA neurons (Fig. 1B). These genes also distinguished SNc from VTA
8 samples from a small cohort (N=3) investigated previously (Nichterwitz et al. 2016) (Fig. 1C). However,
9 few overlapping DEGs were identified between these two studies, even though the same experimental
10 method was used. In fact, only seven and 21 DEGs among the SNc and VTA gene lists, respectively
11 overlapped (Fig. 1D). Notably, the 100 DEGs identified in the small cohort (Nichterwitz et al. 2016)
12 (Supplemental Fig. S2R), failed to distinguish SNc and VTA for the current larger cohort of 18 subjects
13 (Supplemental Fig. S2S; Supplemental Table S7). This suggests that sample size affects identification of
14 DEGs and that a larger cohort size may identify DEGs that more robustly distinguishes the two populations.
15 Moreover, this straight-forward DESeq2 analysis, that considers all samples together, cannot address how
16 significant the 74 DEGs would be when considering smaller patient subsets from the cohort.

17

18 **Bootstrapping coupled with DESeq2 identifies stable DEGs unique to human SNc or VTA**

19 To evaluate how cohort subsets may affect DEG detection, we used a bootstrapping algorithm in
20 combination with DESeq2. Importantly, to reduce the biological variability we considered only those
21 subjects for which both SNc and VTA samples were available (12/18 subjects, 24 samples in total). To
22 begin with this approach, a subset of three subjects were randomly chosen from the pool of total 12 subjects.
23 Differential expression analysis was then performed between the SNc and VTA samples of these subjects
24 (DESeq2), and genes were scored as differentially expressed (DE) or not (adj P-value < 0.05). This random
25 sampling of three subjects, followed by DESeq2 analysis, was performed a total of 1000 times, and the DE
26 frequency over these 1000 comparisons was recorded for this iteration (i = 3 subjects). Subsequently, this
27 process was repeated using subsets of four subjects, then five, up to a maximum of 11 of the 12 subjects.

1 For each subset size (i^3 to i^{11}) the DE frequency was recorded for each gene over the 1000x comparisons of
2 that iteration (Fig. 2A). By considering all DE genes in an iteration (1000x comparisons), we were able to
3 detect hundreds of genes, much more than were detected in any individual comparison, where on average
4 only e.g. 13 DE genes were detected in i^3 (Supplemental Fig. S3A). Interestingly, we found that eight
5 subjects were required to saturate the sensitivity of our gene detection, as few new genes were identified
6 when considering additional subjects in subsequent iterations (Fig. 2B).

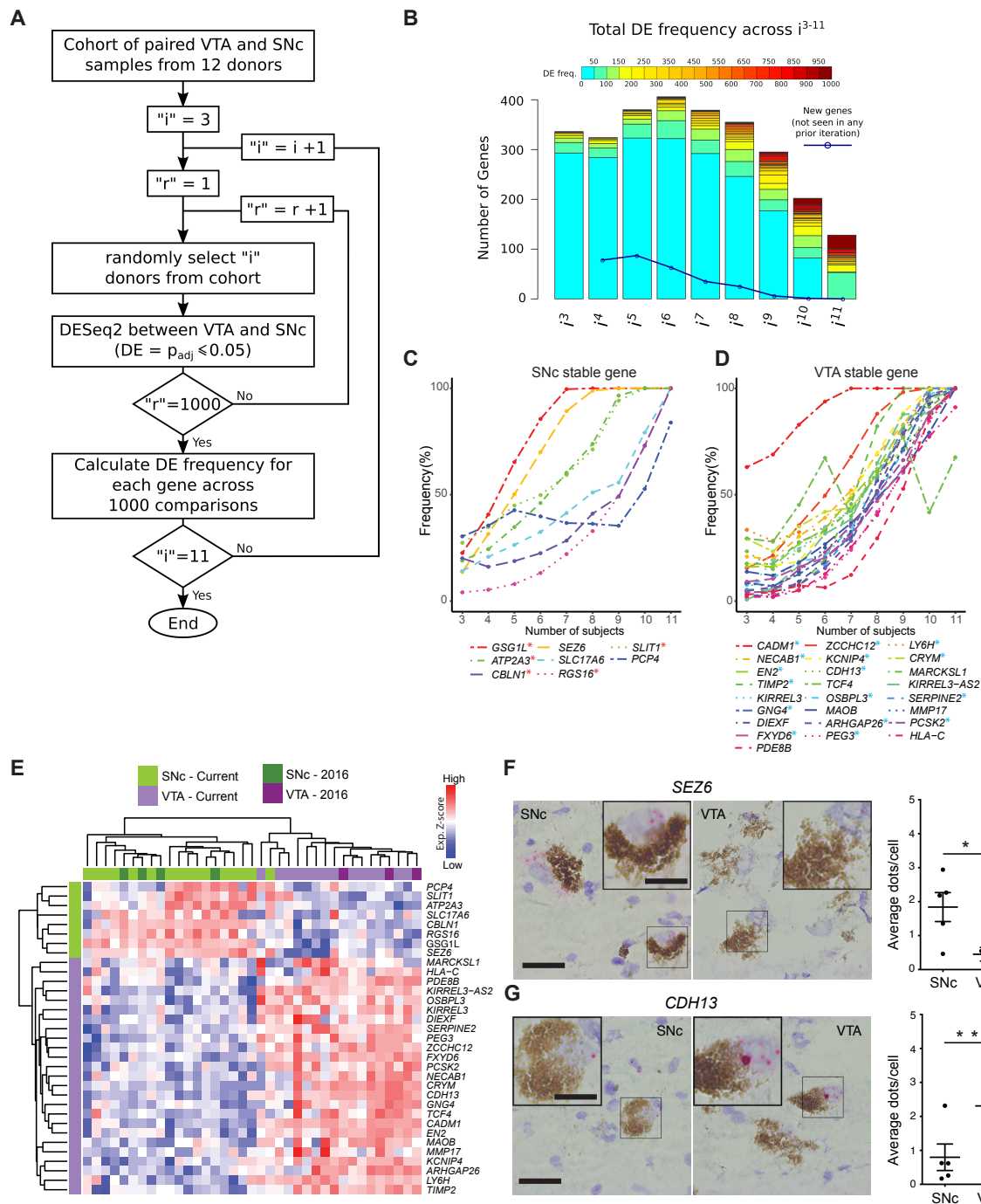


Figure 2

1
 2 **Figure 2.** Bootstrapping analysis coupled with DESeq2 identifies SNc and VTA stable genes from a cohort
 3 of 12 human subjects. (A) Schematic of the analysis workflow. (B) Numbers of genes displaying a given
 4 DE frequency in each iteration of the workflow, from $i = 3$ subjects, up to $i = 11$ subjects. (e.g. in i^3 , almost
 5 300 genes were DE between 1-50 times out of 1000 x r). (C, D) Stable gene DE frequency increased at each

1 iterative sample size increase, for SNc (C) or VTA (D). For each gene list, genes that overlap with DEGs
2 from Figure 1D are marked with an asterisk (*). (E) Hierarchical clustering analysis using the 33 stable
3 genes, faithfully segregated both SNc and VTA samples. (F, G) RNAscope staining and quantification for
4 the stable genes *SEZ6* (F, $p=0.025$) and *CDH13* (G, $p=0.005$) enriched in the SNc and VTA, respectively
5 ($n=5$ subjects, data represented as mean \pm SEM, Paired t test). Scale bars 30 μm (15 μm for insets). See
6 also Supplemental Fig. S3.

7
8 We then summed the DE frequency across i^3 to i^{11} (9000 comparisons in total), separated genes into SNc
9 or VTA enriched lists, and ranked the lists from most to least frequently DEs (Supplemental Fig. S3C, D;
10 Supplemental Table S8). Highly ranked genes on these two lists included multiple known human (e.g.
11 *GSG1L*, *SLIT1*, *ATP2A3*, *CADMI*, *CRYM*, *TCF4*) and rodent rodent (e.g. *Gsg1l*, *Slit2*, *Atp2a3*, *Cadm1*,
12 *Tcf4*) SNc and VTA markers (Grimm et al. 2004; Chung et al. 2005; Greene et al. 2005; Cantuti-Castelvetri
13 et al. 2007; Simunovic et al. 2009; Bifsha et al. 2014; Poulin et al. 2014; La Manno et al. 2016; Nichterwitz
14 et al. 2016). To identify the most reliable DE genes, we designated genes that were detected more than 3000
15 times (out of 9000) as “stable genes”. This stringent cutoff was chosen since the resulting SNc and VTA
16 lists would then contain at least one stable gene that could be identified during the first iteration (where
17 three individuals were used as the sample size).

18
19 We identified eight stable genes for the SNc (Fig. 2C) and 25 stable genes for the VTA (Fig. 2D). Five SNc
20 (labeled in red*: *GSG1L*, *ATP2A3*, *CBLN1*, *RGS16* and *SLIT1*) and 15 VTA (in blue*: *CADMI*, *NECAB1*,
21 *EN2*, *TIMP2*, *GNG4*, *FXVD6*, *ZCCHC12*, *KCNIP4*, *CDH13*, *OSBPL3*, *ARHGAP26*, *PEG3*, *LYH6*, *CRYM*,
22 *SERPINE2* and *PCSK2*) stable genes were among the aforementioned 7 and 21 overlapping DEGs,
23 identified across the two studies (Fig. 1D). We also compared this stable gene list with the outcome of
24 DESeq2 analysis alone, when applied to the same 12 subjects (Supplemental Fig. S3E, F). Reassuringly,
25 all 8 SNc and 25 VTA markers perfectly overlapped with the DEGs from DESeq2 alone using an adjusted
26 P -value <0.05 (Supplemental Fig. S3E) or a more stringent significance (adj. P -value <0.01 , Supplemental
27 Fig. S3F). The expression of SNc stable genes was confirmed in two independent human microarray

1 datasets (which lacked VTA samples) (Supplemental Fig. S3G) (Cantuti-Castelvetri et al. 2007; Simunovic
2 et al. 2009). Further, we used RNAscope (Supplemental Fig. S3H) (Wang et al. 2012) to validate the
3 expression pattern of the SNc stable gene *SEZ6*, and the VTA stable gene *CDH13*, in *post mortem* human
4 tissues (Fig. 2F-G), further ratifying our LCM-seq data and the bootstrapping approach. Importantly, the
5 stable genes faithfully classified SNc and VTA from 21 individuals (Fig. 2E), namely all 18 male
6 individuals from our current dataset and the three female samples investigated previously (Nichterwitz et
7 al. 2016).

8

9 In conclusion, we have identified 33 markers that correctly classify samples as either SNc or VTA, and that
10 remain robust to individual subject variability. Notably, these genes were stably differentially expressed
11 only when at least eight subjects were included in the bootstrapping strategy (Fig. 2D, E), thereby defining
12 a minimal cohort size required to distinguish SNc and VTA samples in human subjects using LCM-seq.

13

14 **Common stable genes can classify SNc and VTA dopamine neurons across species**

15 To identify common stable DEGs for VTA and SNc across other species, we applied the combined
16 bootstrapping and DESeq2 strategy to a single-cell RNA-sequencing (scRNA-seq) dataset that profiled
17 postnatal mouse midbrain dopamine neurons (Supplemental Fig. S4A, raw data analyzed here) (La Manno
18 et al. 2016). Single cells were initially quality controlled for expression of known dopamine neuron
19 markers, and the absence of contaminating glia or oligodendrocytes markers (Supplemental Fig. S4B)
20 (Zhang et al. 2014), followed by iterative bootstrapping and DESeq2 analysis (Supplemental Fig. S4C).

21

22

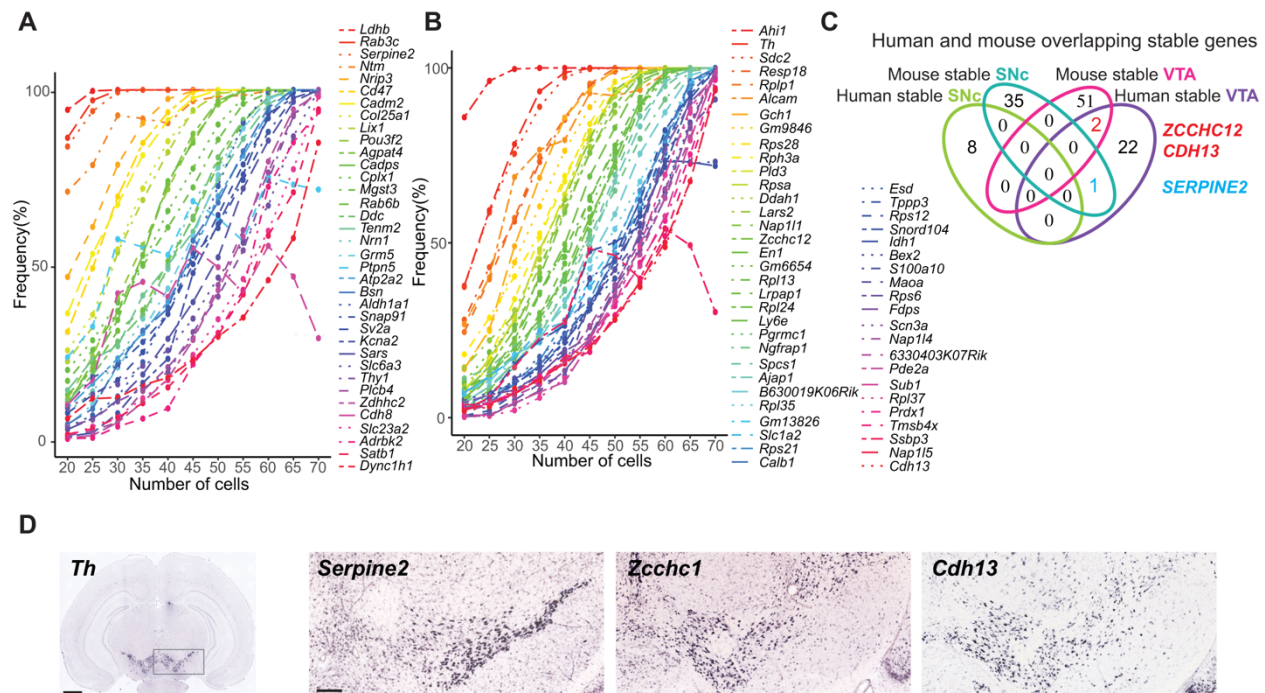


Figure 3

1
2 **Figure 3.** Bootstrapping analysis coupled with DESeq2 identifies SNc or VTA stable genes from a mouse
3 scRNA-seq dataset. (A, B) Identification of SNc (A) and VTA (B) stable genes from 146 single mouse
4 dopamine neurons (La Manno et. al, 2016) (73 available SNc cells and 73 randomly selected VTA cells),
5 see METHODS. (C) Venn-diagram displaying overlapping stable genes between mouse and man. The three
6 common markers *ZCCHC12*, *CDH13* and *SERPINE2* are highlighted in red. (D) *In situ* hybridization
7 images of coronal midbrain sections of the adult mouse (P56, Allen Brain Atlas) showing RNA expression
8 pattern for *Th* and the common stable genes *Serpine2*, *Zcch12* and *Cdh13*. Scale bars: (D) 1mm for *Th* and
9 200 μ m for remaining markers. See also Supplemental Fig. S4.
10
11 This identified 36 SNc- and 53 VTA-specific mouse stable genes (Fig. 3A, B; Supplemental Table S9,
12 Supplemental Fig. S4D, E, red labels) including novel genes (e.g. *Pde2a*, *Adrbk2*, *Tppp3*, *Col25a1*, *Rph3a*,
13 *Mgst3*) in addition to previously reported markers (e.g. *Atp2a2*, *Cadm2*, *Serpine2*, *Zcchc12*, *Ly6e*, *Calb1*)
14 (Grimm et al. 2004; Chung et al. 2005; Greene et al. 2005; Bifsha et al. 2014; Poulin et al. 2014; La Manno
15 et al. 2016; Nichterwitz et al. 2016). Multiple known SNc and VTA markers ranked just below our 30%
16 frequency cutoff and were thus not designated as stable genes (Supplemental Fig. S4D, E, black labels).

1 Importantly, the 89 stable genes correctly classified the mouse single cells as either SNc or VTA
2 (Supplemental Fig. S4F).
3
4 We next cross compared our human and mouse stable gene lists. Two genes, *ZCCHC12* and *CDH13*, were
5 common to human and mouse VTA. In addition, *SERPINE2* was expressed in human VTA, while it's
6 mouse homolog *Serpine2*, was expressed in the mouse SNc (Fig. 3C). The expression patterns of these
7 three genes were corroborated in the adult mouse using Allen *in situ* images (Fig. 3D). To investigate the
8 predictive power of these three common genes, we computationally classified human LCM-seq or mouse
9 scRNA-seq samples using the human and mouse stable gene lists, while including or omitting the three
10 common genes (Fig. 4A-H). (100% reflects perfect classification accuracy, while 50% reflects a random
11 outcome).
12

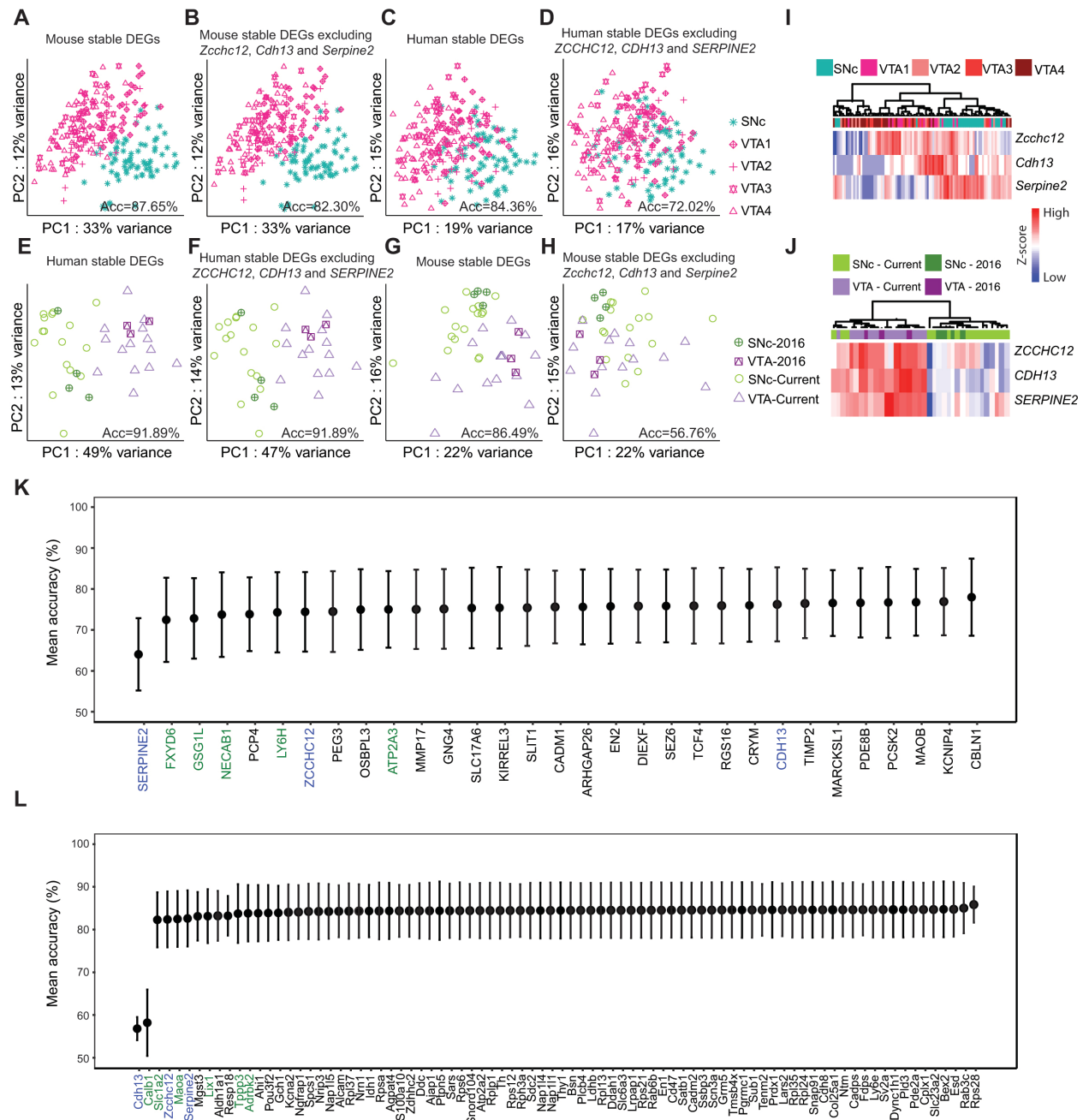


Figure 4

1
 2 **Figure 4.** Predictive power of mouse and human stable genes to resolve SNc and VTA across species. (A-
 3 D) PCAs of the mouse single cells when considering: all mouse stable genes (A), mouse stable genes
 4 excluding the three common genes (B), human stable genes (C) or human stable list excluding the three
 5 common genes (D). (Classification accuracy = Acc%) (E-H) Similar analysis as A-D but applied to 21
 6 human subjects (includes Nichterwitz et al., 2016, subjects). (I, J) Hierarchical clustering analysis
 7 considering only the three common stable genes are sufficient to classify SNc and VTA samples in mouse

1 (I) and human (J). (K) Ranked mean accuracy for individual human stable genes to resolve the mouse
2 dataset, upon random exclusion together with any other two stable genes. The three common stable genes
3 between mouse and human are labelled in blue, while additional genes with predictive value in the mouse
4 are labelled in green. (L) Ranked mean accuracy for individual mouse stable genes to resolve the human
5 dataset, upon random exclusion together with any other two stable genes. Common stable genes between
6 mouse and human are labelled in blue while additional genes with some predictive value in the human (led
7 by *Calb1*) are labelled in green. See also Supplemental Fig. S5.

8

9

10 When using a complete stable gene list to resolve samples within the same species, negligible reduction in
11 the accuracy was observed after removal of the three common genes in either mouse (Fig. 4A, B) or human
12 (Fig. 4E, F). Despite the low overlap of stable genes across mouse and human, they could to some extent
13 separate samples from the other species (Fig. 4C, G). However, subsequent exclusion of the three common
14 genes led to a reduction in classification accuracy (Fig. 4D, H). This was most pronounced when attempting
15 to classify human samples using the mouse stable genes excluding *Zcchc12*, *Cdh13* and *Serpine2*, resulting
16 in an almost chance accuracy of 56.76% (Fig. 4H). Reassuringly, these three markers alone were able to
17 largely segregate SNc from VTA samples in both species (Fig. 4I, J). In conclusion, we identified 89
18 markers for the mouse SNc and VTA using the combined bootstrapping and DESeq2 approach. Cross-
19 species analysis further revealed three mouse stable genes overlapping with the human stable gene list, that
20 alone are sufficient to correctly classify these two neuron subpopulations in both species.

21

22 **Systematic analysis of classification accuracy identifies additional predictive SNc and VTA markers**

23 We next sought to systematically interrogate if the remaining stable human and mouse genes that, despite
24 not overlapping between species, could prove useful in cross-species classification. For a given species, we
25 randomly excluded any three stable genes and calculated the mean accuracy of the remaining markers in
26 resolving the other species' SNc and VTA samples (Fig. 4K, L). Our random exclusion paradigm
27 demonstrated that *SERPINE2* was the most important gene resolving mouse SNc and VTA samples, as it

1 occupied the top rank, while *ZCCHC12* and *CDH13* occupied ranks 7 and 24, respectively (Fig. 4K).
2 Interestingly, the human stable genes *FXYD6*, *GSGIL*, *ATP2A3*, *NECAB1* and *LY6H*, were also highly
3 ranked in their accuracy to classify mouse samples (Fig. 4K). The expression patterns of these genes were
4 corroborated in the adult mouse using Allen *in situ* images (Supplemental Fig. 3D and S5A).
5
6 Unsurprisingly, *Cdh13*, *Zcchc12* and *Serpine2* were top ranked (1, 4 and 6 respectively) upon exclusion
7 from the mouse stable gene list when resolving human samples (Fig. 4L). However, only *Cdh13* strongly
8 contributed to correct sample classification, along with *Calb1*, a known VTA marker in both mouse (Chung
9 et al. 2005; Greene et al. 2005; La Manno et al. 2016) and human (Reyes et al. 2012; Nichterwitz et al.
10 2016). Accordingly, the rank of *CALB1* on the human VTA list was close to the 30% frequency threshold
11 for the “stable gene” classification (Supplemental Fig. S3B; Supplemental Table S8).
12
13 The exclusion of other mouse stable genes, including *Slc1a2*, *Maoa*, *Tppp3*, *Lix1* and *Adrbk2*, which also
14 fell outside of the 30% frequency threshold for human stable genes (Supplemental Table S8), barely
15 impaired the classification of human samples (Fig. 4L). Similar to *Calb1*, *in situ* analysis of these genes in
16 the adult mouse was consistent with their differential expression between VTA and SNc (Supplemental Fig.
17 S5B, green labels). Many novel mouse stable genes, while displaying the expected expression patterns in
18 line with their SNc or VTA enrichment (Supplemental Fig. S5B, black labels), had no impact on the
19 classification of human samples (Fig. 4L). In conclusion, we identified additional human stable genes,
20 including *FXYD6*, *GSGIL*, *ATP2A3*, *NECAB1* and *LY6H* with power to also distinguish SNc or VTA
21 dopamine neurons in mouse.

22

23

24 **DISCUSSION**

25 The selective vulnerability of SNc dopamine neurons to PD, and the relative resilience of VTA dopamine
26 neurons, has encouraged the field to investigate the molecular signature of these two neuron subpopulations.

1 When we analyzed existing data sets (Grimm et al. 2004; Chung et al. 2005; Greene et al. 2005; Bifsha et
2 al. 2014; Poulin et al. 2014; La Manno et al. 2016; Nichterwitz et al. 2016), we identified large discrepancies
3 in the reported SNc or VTA enriched genes across different studies. Such inconsistencies could result from
4 multiple factors, including small sample sizes, as well as variability between subjects, which is recognized
5 to be a major confounding factor in human studies (Mele et al. 2015). This encouraged us to conduct a large
6 focused study on adult human midbrain dopamine neurons using LCM-seq (Nichterwitz et al. 2018). We
7 consequently constructed a comprehensive LCM-seq dataset, isolating single SNc or VTA dopamine
8 neurons from *post mortem* tissues of total 18 individuals, the largest collection of human dopamine neurons,
9 aiming to reveal robust molecular signatures to distinguish the two subpopulations.

10 Using an iterative bootstrapping without replacement coupled with DESeq2 (available at
11 <https://github.com/shanglicheng/BootstrappingWithoutReplacement>), and a strict selection criteria (here, a
12 30% threshold for stable classification) we identify 33 of the most stable DEGs. 25 of these genes define
13 VTA identity, while eight define SNc identity, and together they can accurately classify LCM-seq samples
14 from our previous (3 females), and current (18 males) patient cohorts. Importantly, we also find that a
15 minimum of eight subjects are required to identify these human stable genes using our criteria, and we
16 speculate that identifying lineage-specific markers between any two highly related cellular sub-populations
17 will also require carefully considered, and appropriately large cohort sizes. Such considerations also apply
18 to studies comparing, for example, healthy and diseased dopamine neurons, that may exhibit potentially
19 subtle pathological changes.

20

21 Application of this approach to a mouse single-cell dataset profiling midbrain dopamine neurons (La Manno
22 et al. 2016), identified 89 stable DEGs, among which *Zcchc12*, *Cdh13* and *Serpine2* overlapped with the
23 human stable gene list. Unexpectedly, these genes alone were sufficient to classify VTA and SNc dopamine
24 neurons in both species, which also validated our approach to enrich robust DEGs. These three genes were
25 so far unexplored in the dopamine system. *Serpine2* (Glia-derived nexin) is a serine protease inhibitor which
26 can promote neurite extension by inhibiting thrombin, and which appears downregulated in Alzheimer's

1 disease (Choi et al. 1995). *Serpine2* promotes biogenesis of secretory granule which is required for
2 neuropeptide sorting, processing and secretion (Kim and Loh 2006). *Cdh13* encodes for adhesion protein
3 13, which together with other family members as *Cdh9* and *Cdh15* are linked to neuropsychiatric disorders
4 (Redies et al. 2012). *Cdh13* can regulate neuronal migration and also has an effect on axonal outgrowth as
5 demonstrated in the serotonergic system (Forero et al. 2017). *Zcchc12* on the other hand is a transcriptional
6 co-activator that positively regulates BMP signaling through interaction with Smads and CBP, and appears
7 necessary for basal forebrain cholinergic neuron gene expression (Cho et al. 2008b). When mutated,
8 *Zcchc12* appears to cause X-linked mental retardation (Cho et al. 2008a). It remains to be investigated how
9 dopamine neurons respond to the loss of any of these three factors.

10

11 Remarkably, the species-specific gene lists could also partly classify SNc and VTA dopamine neurons of
12 the other species. Moreover, some stable genes' closely related family members are also expressed in SNc
13 and VTA samples, e.g. *Atp2a2*, *Cadm2*, *Ly6e*, *En1*, *Maoa* and *Pde2a* are present in the mouse while
14 *ATP2A3*, *CADMI*, *LY6H*, *EN2*, *MAOB* and *PDE8B* are found in human. Further, our analysis also
15 demonstrates important differences between the human or mouse stable genes. *Serpine2* was expressed in
16 mouse SNc, while *SERPINE2* was expressed in human VTA. In addition to the three common markers,
17 several human stable genes (e.g. *FXYD6*, *GSGIL*, *ATP2A3*, *NECAB1* and *LY6H*) contribute in classifying
18 mouse SNc and VTA samples. Hence, there remains a solid rational to continue investigating these human-
19 relevant genes in the mouse.

20

21 The stable DEGs we identified here may be highly relevant to induce resistance or model disease as
22 previously attempted in rodents (Chung et al. 2005; Poulin et al. 2014). Several of the human stable genes
23 (or related family members) e.g. *GSGIL*, *ATP2A3*, *SLC17A6*, *RGS16*, *KCNIP1*, *CDH13*, *TCF12*,
24 *OSBPL1A*, *OSBPL10*, *GNG7*, *ARHGAP18*, *ARHGAP24*, *PCSK5*, *PEG3*, *HLA-DOA*, *HLA-DRA*, *HLA-*
25 *DRB1* and *PDE8B* have been found dysregulated in PD (Cantuti-Castelvetri et al. 2007; Bossers et al. 2009;
26 Simunovic et al. 2009) and/or are represented in PD datasets from genome wide association studies

1 (GWASdb SNP-Disease Associations dataset, <http://amp.pharm.mssm.edu>). Interestingly, mice lacking
2 *Rgs6*, a related family member of the human SNc stable gene *RGS16*, develop specific degeneration and
3 cell loss of SNc dopamine neurons at the age of 12 months (Bifsha et al. 2014). Loss of the SNc stable gene
4 *Cplx1* results in a compromised nigrostriatal pathway in knockout mice (Hook et al. 2018). Moreover,
5 mutations in the human SNc stable gene *SEZ6* have been implicated in diseases such as Alzheimer's
6 (Khoonsari et al. 2016; Paracchini et al. 2018), childhood-onset schizophrenia (Ambalavanan et al. 2016),
7 epilepsy and febrile seizures (Yu et al. 2007; Mulley et al. 2011).

8
9 Regarding cell replacement therapies targeting PD (Kriks et al. 2011; Ganat et al. 2012; Hallett et al. 2014;
10 Kefalopoulou et al. 2014; Kirkeby et al. 2017), there remains an urgent need to optimize the pluripotent
11 stem cell preparations to specifically generate SNc rather than VTA neurons (Barker et al. 2017; Sonntag
12 et al. 2018). Evaluation of the correct patterning and differentiation of pluripotent cells to midbrain
13 dopamine neurons relies upon gene expression analysis using quantitative real time PCR (qPCR) or global
14 transcriptome approaches such as RNA sequencing (Ganat et al. 2012; Barker et al. 2017; Nolbrant et al.
15 2017; Studer 2017). Hence, accurate reference gene signatures of adult human SNc neurons are critical
16 towards further advancements in the regenerative PD field. Our LCM-seq and computational stable gene
17 analysis can therefore serve as a base criterion describing the transcriptional profile of adult, human SNc
18 and VTA neurons. This will greatly facilitate dopamine neuron replacement efforts, in addition to disease
19 modeling studies using dopamine neurons derived from patient-specific pluripotent cells (Miller et al. 2013;
20 Vera et al. 2016).

21
22 In summary, using LCM-seq and a bootstrapping approach coupled with DESeq2, we have identified SNc
23 and VTA dopamine neuron markers in both human and mouse. We reveal that the smallest human cohort
24 required to detect such stable DEGs consists of eight individuals, informing future study designs targeting
25 highly related cellular populations. We find that only three genes are equally and strongly predictive across

1 the two species, highlighting the need to carefully define human dopamine neuron subpopulations even
2 when mouse marker profiles are available. This human transcriptomic data set, derived from individually
3 isolated dopamine neurons, will thus help further our understanding and modeling of selective neuronal
4 vulnerability and resilience, and serve as a reference for derivation of authentic SNc or VTA dopamine
5 neurons from stem cells.

6

7 **METHODS**

8

9 **Ethics statement**

10 We have ethical approval to work with human *post mortem* samples (Supplemental Tables S2 and S3) from
11 the regional ethical review board of Stockholm, Sweden (EPN Dnr 2012/111-31/1; 2012/2091-32). Fresh
12 frozen tissue was obtained through the Netherlands Brain Bank (NBB). The work with human tissues was
13 carried out according to the Code of Ethics of the World Medical Association (Declaration of Helsinki).

14

15 **Tissue sectioning and laser capture**

16 Sample preparation prior LCM-seq was carried out as follows. Frozen midbrain tissues obtained from the
17 brain banks were attached to chucks using pre-cooled OCT embedding medium (Histolab). 10 μm -thick
18 coronal sections were acquired in a cryostat at $-20\text{ }^{\circ}\text{C}$ and placed onto precooled-PEN membrane glass
19 slides (Zeiss). For RNAscope experiments, sections were cut at 12 μm -thickness and attached to
20 Superfrost® Plus slides (Thermo Scientific). The slides with sections were kept at $-20\text{ }^{\circ}\text{C}$ during the
21 sectioning and subsequently stored at $-80\text{ }^{\circ}\text{C}$ until further processed.

22 The laser capture procedure followed by sequencing library preparation (LCM-seq) was carried out as
23 described (Nichterwitz et al. 2016; Nichterwitz et al. 2018).

24

25 **Mapping and gene expression quantification**

1 Samples were sequenced using an Illumina HiSeq2000 or HiSeq2500 platform (reads of 43 or 50 bp in
2 length respectively). The uniquely mapped reads were obtained by mapping to the human reference genome
3 hg38/GRCh38 using STAR with default settings. The reads per kilobase of transcript per million mapped
4 reads (RPKM) were estimated using “rpkmforgenes” (Ramskold et al. 2009). As part of the quality control
5 we verified that samples had >1 million reads and >7000 genes expressed with RPKM>1. As all samples
6 processed exceeded this cutoff and had between 1.3~9.8 million reads and 7.9 ~ 12.3 kilo genes expressed
7 with RPKM>1, all samples were included. The correlation coefficient between any two nearest samples
8 was considered to be above 0.7. For cases having more than one replicate per group, corresponding samples
9 were averaged before analysis so that each case had only one SNc and one VTA. Additionally, we
10 confirmed the expression of known midbrain dopamine neuron markers and the purity of each sample (Fig.
11 1A).

12

13 **Differential expression analyses**

14 Differentially expressed genes were identified using the R package “DESeq2” (version: 1.16.1) (Love et al.
15 2014) where the cutoff for significance was an adjusted P value of 0.05. Identified DEGs (from different
16 analysis and summarized below) are shown in Supplemental Tables S4 and S6 to S9.

17 Supplemental Table S4: Mouse differentially expressed genes (La Manno et al. 2016): 390 DEGs calculated
18 from 73 SNc and 170 VTA cells.

19 Supplemental Table S6: Human differentially expressed genes (current study): 74 DEGs calculated from
20 16 SNc and 14 VTA samples from 18 male individuals.

21 Supplemental Table S7: Human differentially expressed genes (Nichterwitz et al. 2016): 100 DEGs
22 calculated from 4 SNc and 3 VTA samples from 3 female subjects.

23 Supplemental Table S8: Human stable genes: 33 DEGs calculated from 12 individuals (from the current
24 study) using the bootstrapping approach.

25 Supplemental Table S9: Mouse stable genes: 89 DEGs calculated from 73 SNc and VTA cells (La Manno
26 et al. 2016) using the bootstrapping approach.

1

2 **Bootstrapping approach coupled with DESeq2**

3 To counteract the variability among human subjects and identify the most reliable DEGs between SNc and
4 VTA neurons across datasets we developed a bootstrapping approach coupled with DESeq2 (Fig. 2A;
5 Supplemental Fig. S4C). The stable genes output of this analysis is correlated with the sample size and give
6 an unbiased estimation of the number of individuals required to consistently distinguish these closely
7 related subpopulations. Importantly this computational tool can be used for the comparison of any other
8 two groups (<https://github.com/shanglicheng/BootstrappingWithoutReplacement>).

9

10 In detail:

11 1) Define “N” and “M” as the number of samples in Group 1 and Group 2, respectively. Choose “I” as a
12 reference representing a given number of samples from “N” and “M”.

13 2) Define “i” as the number of randomly selected samples from Group1 and Group2, where $i \in \{3, 4, 5, \dots,$
14 $I-1\}$. In the human dataset, as we have 12 paired samples, the $i \in \{3, 4, 5, 6, 7, 8, 9, 10, 11\}$.

15 3) Pool “i” samples (temporary considered a “new data set”) and calculate DEGs with DESeq2.

16 4) Repeat steps 2) and 3) for “j” times (set to 1000 times in this study).

17 5) For every round of random selection and DESeq2, save the full list of DEGs, compute and rank their
18 frequency.

19 6) Set a threshold (30% ratio in this study) to consider DEGs with higher frequency as stable genes.

20 Reliable genes appear when frequencies are above: Total times of random pooling x ratio (300 in this study).

21 A stringent, but fair ratio can be defined by comparing the percentage of identified stable genes overlapping
22 with the top (most significant) 10%, 20%, 30%, ..., DEGs identified by DESeq2 alone.

23

24 **Bootstrapping approach applied to human samples**

1 To reliably identify DEGs between human SNc and VTA samples, while minimizing subject variability,
2 we selected 12 individuals (66% of the dataset, 12 out of 18 individuals) where both neuronal populations
3 were available and sequenced. Hence, the number of randomly selected samples (“n” and “m” from “i”
4 individuals) was from three to 11 and the algorithm repeated 1000 times (Fig. 2A).

5

6 **Bootstrapping approach applied to mouse single cells**

7 For this adult mouse dataset (La Mano et. al, 2016) we defined the groups SNc (N=73 cells) and VTA
8 (M=170 cells comprising VTA1, VTA2, VTA3 and VTA4). To compensate the unbalance in cell number
9 and adjust dataset representation compared to the human analysis (66%), we first randomly collected a
10 subset of 73 VTA cells, pairing both SNc and VTA. Similarly, the number of randomly selected samples
11 was 20, 25, 30, ..., 70 and the approach repeated again 1000 times (Supplemental Fig. S4C).

12

13 **Data visualization**

14 Data visualization was achieved using Principal Component Analysis (PCA) and Hierarchical Clustering
15 (H-cluster). PCA was calculated with the function “prcomp” in R with default parameters. Then samples
16 are projected onto the first two dimensions, PC1 and PC2. For H-cluster we used the R function “pheatmap”
17 (version 1.0.12) with the clustering method of “ward.D2”.

18

19 **Accuracy calculation**

20 To measure the power of stable genes in resolving SNc and VTA samples we calculated the clustering
21 accuracy. This is a quantitative parameter to evaluate correct group prediction using hierarchical clustering
22 on the first two dimensions (PC1 and PC2) of the PCA for dimensional reduction. Hierarchical clustering
23 is performed using the ward.D2 method on the selected principal components. Ward criterion is used in the
24 hierarchical clustering because it is based on the multidimensional variance like principal component
25 analysis. The partitioning of samples is based on the first branch of the hierarchical tree. The genes used
26 for PCA are different according to the evaluation purpose.

1 As we have two predicted groups (SNc and VTA), the minimum value for the accuracy is 50%. In details:

2 1) Generate PCA of SNc and VTA.

3 2) Hierarchically cluster the samples by using the first two principle components, PC1 and PC2. Here we
4 can obtain a cluster tree.

5 3) Group the samples into predicted SNc and predicted VTA by the two main branches based on the cluster
6 tree.

7 4) Define "N" and "M" as the total number of samples in Group 1 and Group 2, respectively.

8 5) Define "n" and "m" as the number of correctly predicted samples for Group1 and Group2, respectively.

9 6) Accuracy, expressed as percent, is calculated according to the formula:

10
$$\text{Accuracy (\%)} = ((n + m) / (N + M)) * 100$$

11 We calculated the accuracy for six different conditions (stable gene sets) as follows:

12 a) Mouse stable genes applied to mouse (Fig. 4A) or human (Fig. 4G) samples.

13 b) Mouse stable genes excluding (*Zcchc12*, *Cdh13*, *Serpine2*) applied to mouse (Fig. 4B) or human
14 (Fig. 4H) samples.

15 c) Human stable genes applied to mouse (Fig. 4C) or human (Fig. 4E) samples.

16 d) Human stable genes excluding (*ZCCHC12*, *CDH13*, *SERPINE2*) applied to mouse (Fig. 4D) or
17 human (Fig. 4F) samples.

18 e) Human stable genes excluding three random markers applied to mouse samples (Fig. 4K).

19 f) Mouse stable genes excluding three random markers applied to human samples (Fig- 4L).

20

21 To identify additional (but not common) stable genes from mouse or human with predicted power to the
22 other species, we performed the above analysis e) and f). Here we ranked the mean accuracy for a given
23 stable gene upon random exclusion with any two others.

24

25 **RNAscope staining of human tissues**

1 RNAscope (Wang et al. 2012) was used to verify the expression of one SNc marker (*SEZ6*) and one VTA-
2 preferential gene (*CDH13*) based on the sequencing data. In brief, midbrain sections of human fresh frozen
3 tissue (Supplemental Table S3) were quickly thawed and fixed with fresh PFA (4% in PBS) for 1 hour at
4 4°C. The RNAscope 2.5 HD Assay - RED Kit (Cat. 322360) was used using manufacturer
5 recommendations.

6 To evaluate the procedure in the midbrain tissue (Supplemental Fig. S3D), we first tested a negative control
7 probe against a bacterial gene (Cat. 310043, *dapB-C1*) and a positive control probe against tyrosine
8 hydroxylase (Cat. 441651, *TH-C1*). Once we set up the assay, midbrain sections were stained with *SEZ6*
9 (Cat. 411351-C1) or *CDH13* probes (Cat. 470011-C1). Slides were counterstained with fresh 50% Gill
10 Solution (Cat. GSH132-1L, Sigma-Aldrich) for 2 minutes, washed in water and dried for 15min at 60°C
11 before mounting with Pertex (Cat. 00811, Histolab).

12 For every sample (n=5), we imaged 5-6 random fields within the SNc and VTA regions. On average
13 194.25 ± 43.02 cells were imaged per region and staining. Pictures were made at 40X magnification using
14 the bright-field of a Leica microscope (DM6000/CTR6500 and DFC310 FX camera). After randomization
15 and coding of all the images, the number of dots within melanized cells (dopamine neurons) were counted
16 using ImageJ (version 1.48) and later the average number of dots per cells determined for each region.
17 Investigators performing the quantification were blinded to the sample, target region (SNc and VTA) and
18 probe staining.

19

20 **Statistical analysis**

21 For this study, statistical analyses were performed using "R". For the RNAscope analysis a paired t test
22 (Prism 6, Version 6.0f) was used to compare the mean average dots per cell (for *SEZ6* or *CDH13* staining)
23 between the SNc and VTA. Where applicable, individual statistical tests are detailed in the figure legends
24 where significance is marked by $P < 0.05$. The number of subjects/cells used for each experiment is listed
25 in the figure or figure legends. Results are expressed as mean \pm SD or SEM as specified in the figure legend.

1

2 **DATA ACCESS**

3 All raw and processed sequencing data generated in this study have been submitted to the NCBI Gene
4 Expression Omnibus (GEO; <https://www.ncbi.nlm.nih.gov/geo/>) under accession number GSE114918.
5 Human samples re-analyzed from the Nichterwitz study (Nichterwitz et al. 2016) are under accession
6 number GSE76514 and mouse adult single cells (La Manno et al. 2016) can be accessed at GSE76381.
7 Previous human microarray studies used to verify SNc stable gene expression were accessed from the
8 European repository ArrayExpress (E-MEXP-1416) (Cantuti-Castelvetri et al. 2007) or raw data requested
9 to Dr. Kai C. Sonntag (Simunovic et al. 2009).

10 A processed table with RPKMs values for the full dataset generated in this study can also be found
11 at Mendeley Data under DOI 10.17632/b7nh33pdmg.1

12

13 **ACKNOWLEDGMENTS**

14 We thank Professor Abdel El Manira for critical reading of and insightful comments on the manuscript. We
15 thank all Hedlund laboratory members for fruitful discussions. Human *post mortem* tissues were kindly
16 received from the Netherlands Brain Bank (NBB). This work was funded by grants to E.H. from
17 Parkinsonfonden (795/15; 910/16; 991/17; 1095/18; 1192/19); EU Joint Programme for Neurodegenerative
18 Disease (JPND) (529-2014-7500); Swedish Medical Research Council (Vetenskapsrådet) (2016-02112);
19 and NEURO Sweden; by grants to Q.D. from Swedish Research Council (Vetenskapsrådet) (2014-2870);
20 Svenska Sällskapet för Medicinsk Forskning and Jeansson's Stiftelser; Åke Wiberg, Karolinska Institutets
21 Forskningsstiftelser and Stiftelsen; by grants to J.A. from Karolinska Institutets Forskningsstiftelser,
22 Stiftelsen för ålderssjukdomar (2014-2018) and Åhlén-stiftelsen (mA5 h18; Ärende Nr. 193031 h19). J.A.
23 was supported by a postdoctoral fellowship from the Swedish Society for Medical Research and N.K. by a
24 postdoctoral fellowship from Hjärnfonden, Sweden.

1 *Author contributions:* Conceptualization, E.H., Q.D. and J.A.; Methodology and Investigation, J.A., M.C.,
2 S.C., N.K., Q.D. and E.H.; Software, Formal Analysis and Visualization, S.C., J.A., and N.K.; Writing-
3 Original Draft, J.A., S.C., Q.D. and E.H.; Writing-review and Editing, J.A., S.C., N.K., M.C., Q.D. and
4 E.H.; Supervision and Project Administration, E.H. and Q.D.; Funding Acquisition, E.H., Q.D. and J.A.

5

6 **DISCLOSURE DECLARATION**

7 The authors declare no competing interests.

8

9 **REFERENCES**

- 10 Ambalavanan A, Girard SL, Ahn K, Zhou S, Dionne-Laporte A, Spiegelman D, Bourassa CV, Gauthier J, Hamdan
11 FF, Xiong L et al. 2016. De novo variants in sporadic cases of childhood onset schizophrenia. *Eur J Hum Genet* **24**:
12 944-948.
- 13 Barker RA, Parmar M, Studer L, Takahashi J. 2017. Human Trials of Stem Cell-Derived Dopamine Neurons for
14 Parkinson's Disease: Dawn of a New Era. *Cell Stem Cell* **21**: 569-573.
- 15 Bifsha P, Yang J, Fisher RA, Drouin J. 2014. Rgs6 is required for adult maintenance of dopaminergic neurons in the
16 ventral substantia nigra. *PLoS Genet* **10**: e1004863.
- 17 Bossers K, Meerhoff G, Balesar R, van Dongen JW, Kruse CG, Swaab DF, Verhaagen J. 2009. Analysis of gene
18 expression in Parkinson's disease: possible involvement of neurotrophic support and axon guidance in dopaminergic
19 cell death. *Brain Pathol* **19**: 91-107.
- 20 Cantuti-Castelvetri I, Keller-McGandy C, Bouzou B, Asteris G, Clark TW, Frosch MP, Standaert DG. 2007. Effects
21 of gender on nigral gene expression and parkinson disease. *Neurobiol Dis* **26**: 606-614.
- 22 Cho G, Bhat SS, Gao J, Collins JS, Rogers RC, Simensen RJ, Schwartz CE, Golden JA, Srivastava AK. 2008a.
23 Evidence that SIZN1 is a candidate X-linked mental retardation gene. *Am J Med Genet A* **146A**: 2644-2650.
- 24 Cho G, Lim Y, Zand D, Golden JA. 2008b. Szn1 is a novel protein that functions as a transcriptional coactivator of
25 bone morphogenic protein signaling. *Mol Cell Biol* **28**: 1565-1572.
- 26 Choi BH, Kim RC, Vaughan PJ, Lau A, Van Nostrand WE, Cotman CW, Cunningham DD. 1995. Decreases in
27 protease nexins in Alzheimer's disease brain. *Neurobiol Aging* **16**: 557-562.
- 28 Chung CY, Seo H, Sonntag KC, Brooks A, Lin L, Isacson O. 2005. Cell type-specific gene expression of midbrain
29 dopaminergic neurons reveals molecules involved in their vulnerability and protection. *Hum Mol Genet* **14**: 1709-
30 1725.
- 31 Dahlstroem A, Fuxe K. 1964. Evidence for the Existence of Monoamine-Containing Neurons in the Central Nervous
32 System. I. Demonstration of Monoamines in the Cell Bodies of Brain Stem Neurons. *Acta Physiol Scand Suppl*:
33 SUPPL 232:231-255.

- 1 Damier P, Hirsch EC, Agid Y, Graybiel AM. 1999a. The substantia nigra of the human brain. I. Nigrosomes and the
2 nigral matrix, a compartmental organization based on calbindin D(28K) immunohistochemistry. *Brain* **122 (Pt 8)**:
3 1421-1436.
- 4 Damier P, Hirsch EC, Agid Y, Graybiel AM. 1999b. The substantia nigra of the human brain. II. Patterns of loss of
5 dopamine-containing neurons in Parkinson's disease. *Brain* **122 (Pt 8)**: 1437-1448.
- 6 Di Salvio M, Di Giovannantonio LG, Omodei D, Acampora D, Simeone A. 2010. Otx2 expression is restricted to
7 dopaminergic neurons of the ventral tegmental area in the adult brain. *Int J Dev Biol* **54**: 939-945.
- 8 Forero A, Rivero O, Waldchen S, Ku HP, Kiser DP, Gartner Y, Pennington LS, Waider J, Gaspar P, Jansch C et al.
9 2017. Cadherin-13 Deficiency Increases Dorsal Raphe 5-HT Neuron Density and Prefrontal Cortex Innervation in the
10 Mouse Brain. *Front Cell Neurosci* **11**: 307.
- 11 Ganat YM, Calder EL, Kriks S, Nelander J, Tu EY, Jia F, Battista D, Harrison N, Parmar M, Tomishima MJ et al.
12 2012. Identification of embryonic stem cell-derived midbrain dopaminergic neurons for engraftment. *J Clin Invest*
13 **122**: 2928-2939.
- 14 Greene JG, Dingledine R, Greenamyre JT. 2005. Gene expression profiling of rat midbrain dopamine neurons:
15 implications for selective vulnerability in parkinsonism. *Neurobiol Dis* **18**: 19-31.
- 16 Grimm J, Mueller A, Hefti F, Rosenthal A. 2004. Molecular basis for catecholaminergic neuron diversity. *Proc Natl*
17 *Acad Sci U S A* **101**: 13891-13896.
- 18 Hallett PJ, Cooper O, Sadi D, Robertson H, Mendez I, Isacson O. 2014. Long-term health of dopaminergic neuron
19 transplants in Parkinson's disease patients. *Cell Rep* **7**: 1755-1761.
- 20 Haque NS, LeBlanc CJ, Isacson O. 1997. Differential dissection of the rat E16 ventral mesencephalon and survival
21 and reinnervation of the 6-OHDA-lesioned striatum by a subset of aldehyde dehydrogenase-positive TH neurons. *Cell*
22 *Transplant* **6**: 239-248.
- 23 Hedlund E, Perlmann T. 2009. Neuronal cell replacement in Parkinson's disease. *J Intern Med* **266**: 358-371.
- 24 Hook PW, McClymont SA, Cannon GH, Law WD, Morton AJ, Goff LA, McCallion AS. 2018. Single-Cell RNA-Seq
25 of Mouse Dopaminergic Neurons Informs Candidate Gene Selection for Sporadic Parkinson Disease. *Am J Hum Genet*
26 **102**: 427-446.
- 27 Kefalopoulou Z, Politis M, Piccini P, Mencacci N, Bhatia K, Jahanshahi M, Widner H, Rehncrona S, Brundin P,
28 Bjorklund A et al. 2014. Long-term clinical outcome of fetal cell transplantation for Parkinson disease: two case
29 reports. *JAMA Neurol* **71**: 83-87.
- 30 Khoonsari PE, Haggmark A, Lonnberg M, Mikus M, Kilander L, Lannfelt L, Bergquist J, Ingelsson M, Nilsson P,
31 Kultima K et al. 2016. Analysis of the Cerebrospinal Fluid Proteome in Alzheimer's Disease. *PLoS One* **11**: e0150672.
- 32 Kim T, Loh YP. 2006. Protease nexin-1 promotes secretory granule biogenesis by preventing granule protein
33 degradation. *Mol Biol Cell* **17**: 789-798.
- 34 Kirkeby A, Nolbrant S, Tiklova K, Heuer A, Kee N, Cardoso T, Ottosson DR, Lelos MJ, Rifles P, Dunnett SB et al.
35 2017. Predictive Markers Guide Differentiation to Improve Graft Outcome in Clinical Translation of hESC-Based
36 Therapy for Parkinson's Disease. *Cell Stem Cell* **20**: 135-148.
- 37 Kriks S, Shim JW, Piao J, Ganat YM, Wakeman DR, Xie Z, Carrillo-Reid L, Auyeung G, Antonacci C, Buch A et al.
38 2011. Dopamine neurons derived from human ES cells efficiently engraft in animal models of Parkinson's disease.
39 *Nature* **480**: 547-551.

- 1 La Manno G, Gyllborg D, Codeluppi S, Nishimura K, Salto C, Zeisel A, Borm LE, Stott SRW, Toledo EM,
2 Villaescusa JC et al. 2016. Molecular Diversity of Midbrain Development in Mouse, Human, and Stem Cells. *Cell*
3 **167**: 566-580 e519.
- 4 Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with
5 DESeq2. *Genome Biol* **15**: 550.
- 6 Mele M, Ferreira PG, Reverter F, DeLuca DS, Monlong J, Sammeth M, Young TR, Goldmann JM, Pervouchine DD,
7 Sullivan TJ et al. 2015. Human genomics. The human transcriptome across tissues and individuals. *Science* **348**: 660-
8 665.
- 9 Miller JD, Ganat YM, Kishinevsky S, Bowman RL, Liu B, Tu EY, Mandal PK, Vera E, Shim JW, Kriks S et al. 2013.
10 Human iPSC-based modeling of late-onset disease via progerin-induced aging. *Cell Stem Cell* **13**: 691-705.
- 11 Mulley JC, Iona X, Hodgson B, Heron SE, Berkovic SF, Scheffer IE, Dibbens LM. 2011. The Role of Seizure-Related
12 SEZ6 as a Susceptibility Gene in Febrile Seizures. *Neurol Res Int* **2011**: 917565.
- 13 Nichterwitz S, Benitez JA, Hoogstraaten R, Deng Q, Hedlund E. 2018. LCM-Seq: A Method for Spatial
14 Transcriptomic Profiling Using Laser Capture Microdissection Coupled with PolyA-Based RNA Sequencing.
15 *Methods Mol Biol* **1649**: 95-110.
- 16 Nichterwitz S, Chen G, Aguila Benitez J, Yilmaz M, Storvall H, Cao M, Sandberg R, Deng Q, Hedlund E. 2016. Laser
17 capture microscopy coupled with Smart-seq2 for precise spatial transcriptomic profiling. *Nat Commun* **7**: 12139.
- 18 Nolbrant S, Heuer A, Parmar M, Kirkeby A. 2017. Generation of high-purity human ventral midbrain dopaminergic
19 progenitors for in vitro maturation and intracerebral transplantation. *Nat Protoc* **12**: 1962-1979.
- 20 Panman L, Papathanou M, Laguna A, Oosterveen T, Volakakis N, Acampora D, Kurtzdotter I, Yoshitake T, Kehr J,
21 Joodmardi E et al. 2014. Sox6 and Otx2 control the specification of substantia nigra and ventral tegmental area
22 dopamine neurons. *Cell Rep* **8**: 1018-1025.
- 23 Paracchini L, Beltrame L, Boeri L, Fusco F, Caffarra P, Marchini S, Albani D, Forloni G. 2018. Exome sequencing
24 in an Italian family with Alzheimer's disease points to a role for seizure-related gene 6 (SEZ6) rare variant R615H.
25 *Alzheimers Res Ther* **10**: 106.
- 26 Poulin JF, Zou J, Drouin-Ouellet J, Kim KY, Cicchetti F, Awatramani RB. 2014. Defining midbrain dopaminergic
27 neuron diversity by single-cell gene expression profiling. *Cell Rep* **9**: 930-943.
- 28 Ramskold D, Wang ET, Burge CB, Sandberg R. 2009. An abundance of ubiquitously expressed genes revealed by
29 tissue transcriptome sequence data. *PLoS Comput Biol* **5**: e1000598.
- 30 Redies C, Hertel N, Hubner CA. 2012. Cadherins and neuropsychiatric disorders. *Brain Res* **1470**: 130-144.
- 31 Reyes S, Fu Y, Double K, Thompson L, Kirik D, Paxinos G, Halliday GM. 2012. GIRK2 expression in dopamine
32 neurons of the substantia nigra and ventral tegmental area. *J Comp Neurol* **520**: 2591-2607.
- 33 Schultzberg M, Dunnett SB, Bjorklund A, Stenevi U, Hokfelt T, Dockray GJ, Goldstein M. 1984. Dopamine and
34 cholecystikinin immunoreactive neurons in mesencephalic grafts reinnervating the neostriatum: evidence for selective
35 growth regulation. *Neuroscience* **12**: 17-32.
- 36 Simunovic F, Yi M, Wang Y, Macey L, Brown LT, Krichevsky AM, Andersen SL, Stephens RM, Benes FM, Sonntag
37 KC. 2009. Gene expression profiling of substantia nigra dopamine neurons: further insights into Parkinson's disease
38 pathology. *Brain* **132**: 1795-1809.

- 1 Sonntag KC, Song B, Lee N, Jung JH, Cha Y, Leblanc P, Neff C, Kong SW, Carter BS, Schweitzer J et al. 2018.
2 Pluripotent stem cell-based therapy for Parkinson's disease: Current status and future prospects. *Prog Neurobiol* **168**:
3 1-20.
- 4 Studer L. 2017. Strategies for bringing stem cell-derived dopamine neurons to the clinic-The NYSTEM trial. *Prog*
5 *Brain Res* **230**: 191-212.
- 6 Thompson L, Barraud P, Andersson E, Kirik D, Bjorklund A. 2005. Identification of dopaminergic neurons of nigral
7 and ventral tegmental area subtypes in grafts of fetal ventral mesencephalon based on cell morphology, protein
8 expression, and efferent projections. *J Neurosci* **25**: 6467-6477.
- 9 Vera E, Bosco N, Studer L. 2016. Generating Late-Onset Human iPSC-Based Disease Models by Inducing Neuronal
10 Age-Related Phenotypes through Telomerase Manipulation. *Cell Rep* **17**: 1184-1192.
- 11 Wang F, Flanagan J, Su N, Wang LC, Bui S, Nielson A, Wu X, Vo HT, Ma XJ, Luo Y. 2012. RNAscope: a novel in
12 situ RNA analysis platform for formalin-fixed, paraffin-embedded tissues. *J Mol Diagn* **14**: 22-29.
- 13 Yu ZL, Jiang JM, Wu DH, Xie HJ, Jiang JJ, Zhou L, Peng L, Bao GS. 2007. Febrile seizures are associated with
14 mutation of seizure-related (SEZ) 6, a brain-specific gene. *J Neurosci Res* **85**: 166-172.
- 15 Zhang Y, Chen K, Sloan SA, Bennett ML, Scholze AR, O'Keefe S, Phatnani HP, Guarnieri P, Caneda C, Ruderisch
16 N et al. 2014. An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the
17 cerebral cortex. *J Neurosci* **34**: 11929-11947.
- 18