Spatial transcriptomics and in silico random pooling identify novel markers of vulnerable and resistant dopamine neurons

Julio Aguila[1,#], Shangli Cheng[2,3#], Nigel Kee[1], Ming Cao[1], Qiaolin Deng[2,3*] and Eva Hedlund[1,4]*
[1] Department of Neuroscience, Karolinska Institutet, Stockholm, 171 77, Sweden
[2] Department of Physiology and Pharmacology, Karolinska Institutet, Stockholm, 171 77, Sweden
[3] Center for Molecular Medicine, Karolinska University Hospital, Stockholm, 171 77, Sweden
[4] Lead contact
[#]Equal contribution
[*]Correspondence: eva.hedlund@ki.se (E.H.) or qiaolin.deng@ki.se (Q.D.)

SUMMARY

Defining transcriptional profiles of substantia nigra pars compacta (SNc) and ventral tegmental area (VTA) dopamine neurons (DNs) is critical to understand their differential vulnerability in Parkinson's disease. Reported transcriptional profiles for these DN subpopulations display extensive variability. To resolve this issue we analyze single isolated SNc and VTA DNs from 18 human *post mortem* brains using LCM-seq, providing the largest transcriptome dataset of adult DNs. Using a unique iterative random pooling algorithm, with high utility to resolve differences across any two subpopulations, we reveal 33 robust DN subtype-specific markers. We define that eight subjects is the minimal cohort size required to identify these stably differentially expressed genes, explaining inconsistent results from smaller cohorts. Finally, we identify three genes, *ZCCHC12*, *CDH13* and *SERPINE2*, that alone can classify SNc or VTA DNs across species. The markers identified will be vital for future studies aiming to induce DN resilience or model disease.

KEYWORDS
Human midbrain dopamine neurons, spatial transcriptomics, laser capture, RNA sequencing, substantia nigra compacta, ventral tegmental area, Parkinson's disease

INTRODUCTION

Midbrain dopamine neurons are divided into two major subpopulations, the substantia nigra pars compacta (SNc) and the ventral tegmental area (VTA) (Hedlund and Perlmann, 2009). SNc dopamine neurons project to the dorsolateral striatum (Dahlstroem and Fuxe, 1964) and are severely affected in Parkinson's Disease (PD)(Damier et al., 1999b), while VTA dopamine neurons project to cortical and mesolimbic areas and are more resilient to degeneration (Hedlund and Perlmann, 2009). These subtypes have been extensively investigated in numerous animal

models towards the goal of identifying molecular mechanisms that can prevent degeneration or to model disease (Bifsha et al., 2014; Chung et al., 2005; Greene et al., 2005; Grimm et al., 2004; Poulin et al., 2014). However, the reported transcriptional profiles for SNc and VTA neurons from those studies display extensive variability and result in very few reproducible markers. As SNc dopamine neurons represent the ideal cell type for transplantation in PD (Ganat et al., 2012; Haque et al., 1997; Kriks et al., 2011; Schultzberg et al., 1984; Thompson et al., 2005), a precise and thorough transcriptional profiling in adult human tissues can serve as a foundation for their *in vitro* derivation.

Here we use LCM-seq to analyze the transcriptome of individually isolated SNc and VTA dopamine neurons from 18 human *post mortem* brains, combined with a unique iterative random pooling algorithm to reveal 33 robust subtype-specific markers. We show that the minimal cohort size required to reliably identify these SNc- and VTA-specific markers were eight subjects, clarifying why smaller cohorts have given inconsistent results. We further validate the algorithm on a large-scale data set on mouse single midbrain dopamine neurons, and reveal three genes, *ZCCHC12*, *CDH13* and *SERPINE2*, which alone can accurately classify SNc or VTA samples in either species.

RESULTS

Bioinformatics analysis of dopamine neuron subpopulations shows inconsistencies across published studies

When we revisit previous publications defining the identity of VTA and SNc midbrain dopamine neurons in mouse and rat (Bifsha et al., 2014; Chung et al., 2005; Greene et al., 2005; Grimm et al., 2004; La Manno et al., 2016), we find that a surprisingly low fraction of DEGs are common across studies (Figures S1A and S1B; Tables S1 and S4). Cross-comparison of these genes with our previously published human dataset (Nichterwitz et al., 2016), identify overlap of only two genes, *SOX6* and *CALB1*, within SNc and VTA gene lists, respectively (Figure S1C). These discrepancies highlight the urgent need to reproducibly and robustly identify signatures for VTA and SNc dopamine neurons, particularly applicable to humans.

Random pooling analysis identify stable genes unique to the human SNc or VTA

We previously applied the highly sensitive spatial transcriptomic method LCM-seq (Nichterwitz et al., 2018; Nichterwitz et al., 2016) to *post mortem* SNc and VTA midbrain dopamine neurons originating from three human subjects. We have now expanded our analysis of individually isolated dopamine neurons to include an additional 18 adult human subjects (Figures S2A-S2H; Table S2). This study represents the largest human dataset profiling SNc and VTA dopamine neurons to date. Differential expression analysis, considering these 18 subjects, identified 74 DEGs (Table S5), which resolved SNc from VTA neurons (Figure 1A). These genes also

distinguished SNc from VTA samples used in our previous study (Nichterwitz et al., 2016) (Figure 1B), but few overlapping DEGs were identified across the two studies. In fact, only seven and 21 DEGs for SNc and VTA gene lists, respectively overlapped (Figure 1C). Notably, the 100 DEGs identified in the small (N=3) cohort (Nichterwitz et al., 2016), which fully resolved SNc and VTA samples there (Figures S2I and S2J; Table S6), failed to do so for the new larger cohort of 18 subjects (Figures S2I and S2J). These findings highlight that the size of and variability among cohorts will affect the identification of lineage-specific markers, resulting in few commonly shared genes across groups and studies.

To address this problem, we designed a random pooling algorithm (Figure 1D, see STAR methods), that iteratively samples an increasing proportion of a cohort of subjects while performing differential expression analysis (DESeq2) at each iteration. Upon completion, genes are ranked according to the cumulative frequency of all iterative differential expression results, as a function of sample size. To reduce the individual variability, we selected 12 subjects from which both SNc and VTA LCM-seq samples were obtained. Application of the random pooling strategy to these 24 samples yielded frequency ranks of DEGs for SNc and VTA (Figures S2K and S2L; Table S7) containing multiple known human and/or rodent SNc and VTA markers (Bifsha et al., 2014; Cantuti-Castelvetri et al., 2007; Chung et al., 2005; Greene et al., 2005; Grimm et al., 2004; La Manno et al., 2016; Nichterwitz et al., 2016; Poulin et al., 2014; Simunovic et al., 2009) (Table S4), thus validating our approach. To identify the most reliable markers we designated genes within the top 30% of frequency scores, as “stable genes”. This approach identified eight stable genes for the SNc (Figure 1E; Table S7) and 25 stable genes for the VTA (Figure 1F; Table S7). Many of these stable genes (labeled in red***** for SNc and blue***** for VTA) were among the aforementioned overlapping DEGs (Figure 1C). Notably, these genes were stably differentially expressed only when at least eight subjects were included in the random pooling strategy (Figures 1E and 1F), suggesting this to be the minimal cohort size required to reproducibly distinguish these two closely related dopamine neuron subpopulations. The expression of the SNc stable genes was confirmed in two independent human microarray datasets (which lacked VTA samples) (Figure S2M) (Cantuti-Castelvetri et al., 2007; Simunovic et al., 2009). Importantly, the 33 stable genes faithfully classified subtypes from 21 subjects (Figure 1G), including all 18 male subjects from our current dataset and the three female subjects investigated previously (Nichterwitz et al., 2016). Using RNAscope (Figure S2N) (Wang et al., 2012) we validated the expression pattern of the SNc stable gene *SEZ6* (Figure 1H) and the VTA stable gene *CDH13* (Figure 1I), further ratifying our LCM-seq data and the random pooling algorithm.

A few stable genes define SNc and VTA dopamine neurons across species

To extend the utility and further validate our approach, we applied our random pooling algorithm to a postnatal mouse single-cell dataset profiling midbrain dopamine neurons (Figures S3A-S3C) (La Manno et al., 2016). Single cells were initially assessed for expression of known dopamine

neuron markers (Figure S3B). Following the random pooling analysis (Figure S3C) we identified 36 stable SNc genes and 53 stable VTA genes (Figures 2A and 2B; Table S8). These 89 genes effectively classified all single cells into the correct subtypes (Figure S3D). The stable gene sets for SNc and VTA (Figures S3E and S3F, red labels) included novel genes in addition to previously reported subtype-specific genes (Bifsha et al., 2014; Chung et al., 2005; Greene et al., 2005; Grimm et al., 2004; La Manno et al., 2016; Nichterwitz et al., 2016; Poulin et al., 2014). Multiple known DEGs ranked just below our 30% frequency cutoff and were thus not designated as stable genes (Figures S3E and S3F, black labels).

We next cross compared our human and mouse stable genes. Two genes, *ZCCHC12* and *CDH13*, were common to human and mouse VTA. In addition, *SERPINE2* was expressed in human VTA, while it's mouse homolog *Serpine2*, was expressed in the mouse SNc (Figure 2C). To investigate the predictive power of these three common genes, we assessed the accuracy of our sample clustering (100% reflects perfect accuracy, while 50% reflects a random outcome, see STAR methods). We assessed both human and mouse samples using the stable genes for each species with or without the three common genes (Figures 2D-2K). As expected, when using a complete stable gene list to resolve samples within the same species, negligible reduction in the accuracy was observed after removal of the three common genes in either mouse (Figures 2D and 2E) or human (Figures 2H and 2I). Despite the poor DEGs overlap across species, both species' stable genes could moderately separate the other species (Figures 2F and 2J). However, subsequent exclusion of the three common genes led to a reduction in classification accuracy (Figures 2G and 2K). This was most pronounced when attempting to classify human samples using the mouse stable genes excluding *Zcchc12*, *Cdh13* and *Serpine2*, resulting in an almost chance accuracy of 56.76% (Figure 2K). These results highlight the significant predictive power of these three novel common genes, which alone can accurately classify SNc or VTA samples in either species (Figures 2L and 2M).

Identification and validation of genes with cross-species predictive power to resolve SNc and VTA dopamine neurons

We next sought to systematically interrogate the remaining stable genes that, despite not overlapping, could prove usefulness in cross-species classification. For a given species, we randomly excluded any three stable genes and calculated the mean accuracy of the remaining markers in resolving the other species' SNc and VTA samples (see STAR methods, Figure 3A; Figure S4A). Exclusion of *CDH13*, *ZCCHC12* and *SERPINE2* from the human stable gene list moderately reduced the accuracy of classifying mouse samples (Figure 2F and 2G). Our random exclusion paradigm demonstrated that *SERPINE2* is the most important gene driving this observation, as it occupied the top rank, while *ZCCHC12* and *CDH13* occupied ranks 7 and 24, respectively (Figure 3A). Further, the human stable genes *FXYD6*, *GSG1L*, *ATP2A3*, *NECAB1* and *LY6H*, were also highly ranked in their accuracy to classify mouse samples (Figure 3A). The expression patterns of these genes in mouse were corroborated using Allen *in situ* images (Figure 3C).

In line with previous results (Figures 2J and 2K), *Cdh13*, *Zcchc12* and *Serpine2* were top ranked (1, 4 and 6 respectively, Figure S4A) upon exclusion from the mouse stable gene list when resolving human samples. These genes also displayed definitive expression patterns in the adult mouse (Figure 3B; Figure S4B, P56 mice, Allen Institute for Brain Science, http://developingmouse.brain-map.org).

Moreover, *Calb1*, a known VTA marker in both mouse (Figure 3C) (Chung et al., 2005; Greene et al., 2005; La Manno et al., 2016) and human (Nichterwitz et al., 2016; Reyes et al., 2012), ranked second highest in resolving human samples (Figure S4A). Accordingly, its rank on the human VTA list was close to the 30% frequency threshold for the "stable gene" classification (Figure S2L). The exclusion of other mouse stable genes, including *Slc1a2*, *Maoa*, *Tppp3*, *Lix1* and *Adrbk2*, which also fell outside of the 30% frequency threshold for human stable genes (Table S7), weakly impaired classification of human samples (Figure S4A). Similar to *Calb1*, *in situ* analysis of these genes in the adult mouse was consistent with their differential expression between VTA and SNc (Figure S4C, green labels).
Many novel mouse stable genes, while displaying the expected expression patterns in line with their SNc or VTA enrichment (Figure S4C, black labels), had little impact on the classification of human samples.

Finally, the low fraction of common mouse and human stable genes (2.5%, 3 common out of 122 genes, Figure 2C) could simply be due to genes falling outside the strict 30% frequency threshold. However, a low overlap still remained (2.8%, 61 common out of 2188 genes) when considering the full outcome of the random pooling algorithm applied to either species (Tables S7 and S8).

DISCUSSION

Understanding the unique traits of SNc and VTA dopamine neurons is paramount to unravelling cell intrinsic mechanisms of vulnerability and resilience in PD. Targeted analysis of these midbrain dopamine neuron subpopulations have revealed several markers that differentially label SNc e.g. *Aldh1a7*, *Sox6*, *Cbln1*, *Vav3*, *Atp2a3* and VTA e.g. *Calb1*, *Otx2*, *Crym*, *Cadm1*, *Marcks* (Bifsha et al., 2014; Chung et al., 2005; Damier et al., 1999a; Di Salvio et al., 2010; Greene et al., 2005; Grimm et al., 2004; Nichterwitz et al., 2016; Panman et al., 2014) dopamine neurons.

Transcriptome studies aimed at conclusively defining these two subpopulations have mainly been conducted in rodent tissues (Bifsha et al., 2014; Bulteau et al., 2017; Chung et al., 2005; Fico et al., 2014; Greene et al., 2005; Grimm et al., 2004; Poulin et al., 2014). However, when cross-referencing these we found large discrepancies across studies that warranted further investigation. Furthermore, as there may be important species differences it is vital to carefully

investigate human neurons and compare with those of rodents. When we cross-compared the few existing transcriptome studies on human dopamine neurons; two studies on SNc only (Cantuti-Castelvetri et al., 2007; Simunovic et al., 2009) and one previous study on VTA and SNc neurons on a small patient cohort (Nichterwitz et al., 2016), with rodent data, only *Sox6* and *Calb1* consistently defined these subpopulations across species. The inconsistencies among rodent studies and across human studies is likely explained by multiple factors including sample size, gender, age and the methods used to profile the cells. These discrepancies encouraged us to conduct a focused study on adult human midbrain neurons using LCM-seq, a refined spatial transcriptomic methodology previously developed in our lab (Nichterwitz et al., 2018; Nichterwitz et al., 2016). We conducted our study on male samples only and collected dopamine neurons from a large cohort consisting of 18 subjects, representing the largest human dataset profiling SNc and VTA dopamine neurons to date.

When comparing DEGs between SNc and VTA across our human studies, we found a rather low overlap between the current dataset of 18 male subjects and that of our previous study including dopamine neurons isolated from only three female subjects, even when the same methodology (LCM-seq) and computational tool (Nichterwitz et al., 2016) was employed. This is not surprising as the one major confounding factor in human studies is the variability between subjects, largely due to ethnicity, gender and age (Mele et al., 2015).

We wanted to further evaluate how sample size would affect the outcome of transcriptomic studies. To do so, we developed a random pooling algorithm that can be used to identify the most reliable DEGs between any two populations as a function of sample size (https://github.com/shanglicheng/RandomPooling). Using this computational tool, with a strict 30% cut off, we identified 33 accurate markers that were able to distinguish SNc and VTA samples from 21 subjects comprising both datasets. Importantly, this analysis showed that at least eight subjects are required to consistently distinguish these closely related subpopulations; a key variable for valid study designs using human samples (Ching et al., 2014). We speculate that any other two closely related subpopulations will require similar cohort sizes to resolve subtle differences.

Our application of the random pooling algorithm to mouse dopamine neurons demonstrate that there are large differences in stable genes across species. In fact, there were only three stable genes (*ZCCHC12*, *CDH13* and *SERPINE2*) that overlapped, but were sufficient to largely classify samples within the same species. Nonetheless, human and mouse specific-stable genes could partly resolve SNc and VTA dopamine neurons of the other species, thus demonstrating general similar features. Interestingly, in addition to the three common markers, several human stable genes (ex. *FXYD6*, *GSG1L*, *NECAB1*, *LY6H* and *ATP2A3*) contribute to distinguishing mouse SNc and VTA samples, and thus there remains a solid rationale to continue investigating these human-relevant genes in the mouse.

The stably differentially expressed genes we identified here may be highly relevant to induce resistance or model disease as previously attempted in rodents (Bifsha et al., 2014; Chung et al., 2005; Poulin et al., 2014).
Transplantation of dopamine neurons derived from human pluripotent sources or fetal ventral midbrain tissue have demonstrated clinical relevance in disease models and patients (Ganat et al., 2012; Hallett et al., 2014; Kefalopoulou et al., 2014; Kirkeby et al., 2017; Kriks et al., 2011). However, there is still an urgent need to optimize the therapeutically relevant cell preparation to specifically generate SNc rather than VTA neurons for PD (Barker et al., 2017; Sonntag et al., 2018). Hence, having the genuine signature of adult human SNc neurons as a reference is crucial towards further advancements in the regenerative PD field. This reference is equally important when deriving midbrain dopamine neurons *in vitro* from patient-specific iPS, not only with the value of providing a specific cell identity but also the maturation properties required for meaningful disease modelling (Miller et al., 2013; Vera et al., 2016).

In summary, we have used spatial transcriptomics combined with a novel random pooling algorithm that enabled us to derive robust and unique lists of SNc and VTA markers for both human and mouse. We reveal that the smallest cohort required to detect stable DEGs between SNc and VTA consists of eight individuals, clarifying the large discrepancies between smaller published cohort, and giving directive for future studies examining cellular populations of high similarity. Interestingly, we find that only three genes are equally and strongly predictive across the two species. This highlights the vital need to appreciate the human marker profiles of these clinically relevant and highly similar dopaminergic sub-types. Such profiles will be invaluable for *in vitro* modeling of selective vulnerability, as well as for the derivation of authentic SNc neurons for use in cell transplantation strategies targeting PD.

ACKNOWLEDGMENTS

This work was funded by grants to E.H. from Parkinsonfonden (795/15; 910/16; 991/17; 1095/18); EU Joint Programme for Neurodegenerative Disease (JPND) (529-2014-7500); Swedish Medical Research Council (Vetenskapsrådet) (2016-02112); and NEURO Sweden; by grants to Q.D. from Swedish Research Council (Vetenskapsrådet) (2014-2870); Svenska Sällskapet för Medicinsk Forskning and Jeanssons Stiftelser; Åke Wiberg, Karolinska Institutets Forskningsstiftelser and Stiftelsen; by grants to J.A. from Karolinska Institutets Forskningsstiftelser and Stiftelsen för ålderssjukdomar (2014-2018). J.A. is supported by a postdoctoral fellowship from the Swedish Society for Medical Research and N.K. by a postdoctoral fellowship from Hjärnfonden, Sweden. Human post-mortem tissues were kindly received from the Netherlands Brain Bank (NBB).

AUTHOR CONTRIBUTIONS

Conceptualization, E.H., Q.D. and J.A.; Methodology and Investigation, J.A., M.C., S.C., N.K., Q.D. and E.H.; Software, Formal Analysis and Visualization, S.C., J.A., and N.K.; Writing-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 E.H.; Supervision and Project Administration, E.H. and Q.D.; Funding Acquisition, E.H., Q.D. and J.A.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

Barker, R.A., Parmar, M., Studer, L., and Takahashi, J. (2017). Human Trials of Stem Cell-Derived Dopamine Neurons for Parkinson's Disease: Dawn of a New Era. Cell Stem Cell *21*, 569-573.

Bifsha, P., Yang, J., Fisher, R.A., and Drouin, J. (2014). Rgs6 is required for adult maintenance of dopaminergic neurons in the ventral substantia nigra. PLoS Genet *10*, e1004863.

Bulteau, A.L., Mena, N.P., Auchere, F., Lee, I., Prigent, A., Lobsiger, C.S., Camadro, J.M., and Hirsch, E.C. (2017). Dysfunction of mitochondrial Lon protease and identification of oxidized protein in mouse brain following exposure to MPTP: Implications for Parkinson disease. Free Radic Biol Med *108*, 236-246.

Cantuti-Castelvetri, I., Keller-McGandy, C., Bouzou, B., Asteris, G., Clark, T.W., Frosch, M.P., and Standaert, D.G. (2007). Effects of gender on nigral gene expression and parkinson disease. Neurobiol Dis *26*, 606-614.

Ching, T., Huang, S., and Garmire, L.X. (2014). Power analysis and sample size estimation for RNA-Seq differential expression. RNA *20*, 1684-1696.

Chung, C.Y., Seo, H., Sonntag, K.C., Brooks, A., Lin, L., and Isacson, O. (2005). Cell type-specific gene expression of midbrain dopaminergic neurons reveals molecules involved in their vulnerability and protection. Hum Mol Genet *14*, 1709-1725.

Dahlstroem, A., and Fuxe, K. (1964). Evidence for the Existence of Monoamine-Containing Neurons in the Central Nervous System. I. Demonstration of Monoamines in the Cell Bodies of Brain Stem Neurons. Acta Physiol Scand Suppl, SUPPL 232:231-255.

Damier, P., Hirsch, E.C., Agid, Y., and Graybiel, A.M. (1999a). The substantia nigra of the human brain. I. Nigrosomes and the nigral matrix, a compartmental organization based on calbindin D(28K) immunohistochemistry. Brain *122 (Pt 8)*, 1421-1436.

Damier, P., Hirsch, E.C., Agid, Y., and Graybiel, A.M. (1999b). The substantia nigra of the human brain. II. Patterns of loss of dopamine-containing neurons in Parkinson's disease. Brain *122 (Pt 8)*, 1437-1448.

Di Salvio, M., Di Giovannantonio, L.G., Omodei, D., Acampora, D., and Simeone, A. (2010). Otx2 expression is restricted to dopaminergic neurons of the ventral tegmental area in the adult brain. Int J Dev Biol *54*, 939-945.

Fico, A., de Chevigny, A., Melon, C., Bohic, M., Kerkerian-Le Goff, L., Maina, F., Dono, R., and Cremer, H. (2014). Reducing glypican-4 in ES cells improves recovery in a rat model of Parkinson's disease by increasing the production of dopaminergic neurons and decreasing teratoma formation. J Neurosci *34*, 8318-8323.

Ganat, Y.M., Calder, E.L., Kriks, S., Nelander, J., Tu, E.Y., Jia, F., Battista, D., Harrison, N., Parmar, M., Tomishima, M.J., *et al.* (2012). Identification of embryonic stem cell-derived midbrain dopaminergic neurons for engraftment. J Clin Invest *122*, 2928-2939.

Greene, J.G., Dingledine, R., and Greenamyre, J.T. (2005). Gene expression profiling of rat midbrain dopamine neurons: implications for selective vulnerability in parkinsonism. Neurobiol Dis *18*, 19-31.

Grimm, J., Mueller, A., Hefti, F., and Rosenthal, A. (2004). Molecular basis for catecholaminergic neuron diversity. Proc Natl Acad Sci U S A *101*, 13891-13896.

Hallett, P.J., Cooper, O., Sadi, D., Robertson, H., Mendez, I., and Isacson, O. (2014). Long-term health of dopaminergic neuron transplants in Parkinson's disease patients. Cell Rep *7*, 1755-1761.

Haque, N.S., LeBlanc, C.J., and Isacson, O. (1997). Differential dissection of the rat E16 ventral mesencephalon and survival and reinnervation of the 6-OHDA-lesioned striatum by a subset of aldehyde dehydrogenase-positive TH neurons. Cell Transplant *6*, 239-248.
Hedlund, E., and Perlmann, T. (2009). Neuronal cell replacement in Parkinson's disease. J Intern Med *266*, 358-371.
Kefalopoulou, Z., Politis, M., Piccini, P., Mencacci, N., Bhatia, K., Jahanshahi, M., Widner, H., Rehncrona, S., Brundin, P., Bjorklund, A., *et al.* (2014). Long-term clinical outcome of fetal cell transplantation for Parkinson disease: two case reports. JAMA Neurol *71*, 83-87.
Kirkeby, A., Nolbrant, S., Tiklova, K., Heuer, A., Kee, N., Cardoso, T., Ottosson, D.R., Lelos, M.J., Rifes, P., Dunnett, S.B., *et al.* (2017). Predictive Markers Guide Differentiation to Improve Graft Outcome in Clinical Translation of hESC-Based Therapy for Parkinson's Disease. Cell Stem Cell *20*, 135-148.
Kriks, S., Shim, J.W., Piao, J., Ganat, Y.M., Wakeman, D.R., Xie, Z., Carrillo-Reid, L., Auyeung, G., Antonacci, C., Buch, A., *et al.* (2011). Dopamine neurons derived from human ES cells efficiently engraft in animal models of Parkinson's disease. Nature *480*, 547-551.
La Manno, G., Gyllborg, D., Codeluppi, S., Nishimura, K., Salto, C., Zeisel, A., Borm, L.E., Stott, S.R.W., Toledo, E.M., Villaescusa, J.C., *et al.* (2016). Molecular Diversity of Midbrain Development in Mouse, Human, and Stem Cells. Cell *167*, 566-580 e519.
Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol *15*, 550.
Mele, M., Ferreira, P.G., Reverter, F., DeLuca, D.S., Monlong, J., Sammeth, M., Young, T.R., Goldmann, J.M., Pervouchine, D.D., Sullivan, T.J., *et al.* (2015). Human genomics. The human transcriptome across tissues and individuals. Science *348*, 660-665.
Miller, J.D., Ganat, Y.M., Kishinevsky, S., Bowman, R.L., Liu, B., Tu, E.Y., Mandal, P.K., Vera, E., Shim, J.W., Kriks, S., *et al.* (2013). Human iPSC-based modeling of late-onset disease via progerin-induced aging. Cell Stem Cell *13*, 691-705.
Nichterwitz, S., Benitez, J.A., Hoogstraaten, R., Deng, Q., and Hedlund, E. (2018). LCM-Seq: A Method for Spatial Transcriptomic Profiling Using Laser Capture Microdissection Coupled with PolyA-Based RNA Sequencing. Methods Mol Biol *1649*, 95-110.
Nichterwitz, S., Chen, G., Aguila Benitez, J., Yilmaz, M., Storvall, H., Cao, M., Sandberg, R., Deng, Q., and Hedlund, E. (2016). Laser capture microscopy coupled with Smart-seq2 for precise spatial transcriptomic profiling. Nat Commun *7*, 12139.
Panman, L., Papathanou, M., Laguna, A., Oosterveen, T., Volakakis, N., Acampora, D., Kurtsdotter, I., Yoshitake, T., Kehr, J., Joodmardi, E., *et al.* (2014). Sox6 and Otx2 control the specification of substantia nigra and ventral tegmental area dopamine neurons. Cell Rep *8*, 1018-1025.
Poulin, J.F., Zou, J., Drouin-Ouellet, J., Kim, K.Y., Cicchetti, F., and Awatramani, R.B. (2014). Defining midbrain dopaminergic neuron diversity by single-cell gene expression profiling. Cell Rep *9*, 930-943.
Ramskold, D., Wang, E.T., Burge, C.B., and Sandberg, R. (2009). An abundance of ubiquitously expressed genes revealed by tissue transcriptome sequence data. PLoS Comput Biol *5*, e1000598.
Reyes, S., Fu, Y., Double, K., Thompson, L., Kirik, D., Paxinos, G., and Halliday, G.M. (2012). GIRK2 expression in dopamine neurons of the substantia nigra and ventral tegmental area. J Comp Neurol *520*, 2591-2607.
Schultzberg, M., Dunnett, S.B., Bjorklund, A., Stenevi, U., Hokfelt, T., Dockray, G.J., and Goldstein, M. (1984). Dopamine and cholecystokinin immunoreactive neurons in mesencephalic grafts reinnervating the neostriatum: evidence for selective growth regulation. Neuroscience *12*, 17-32.
Simunovic, F., Yi, M., Wang, Y., Macey, L., Brown, L.T., Krichevsky, A.M., Andersen, S.L., Stephens, R.M., Benes, F.M., and Sonntag, K.C. (2009). Gene expression profiling of substantia nigra dopamine neurons: further insights into Parkinson's disease pathology. Brain *132*, 1795-1809.
Sonntag, K.C., Song, B., Lee, N., Jung, J.H., Cha, Y., Leblanc, P., Neff, C., Kong, S.W., Carter, B.S., Schweitzer, J., *et al.* (2018). Pluripotent stem cell-based therapy for Parkinson's disease: Current status and future prospects. Prog Neurobiol *168*, 1-20.
Thompson, L., Barraud, P., Andersson, E., Kirik, D., and Bjorklund, A. (2005). Identification of dopaminergic neurons of nigral and ventral tegmental area subtypes in grafts of fetal ventral mesencephalon based on cell morphology, protein expression, and efferent projections. J Neurosci *25*, 6467-6477.
Vera, E., Bosco, N., and Studer, L. (2016). Generating Late-Onset Human iPSC-Based Disease Models by Inducing Neuronal Age-Related Phenotypes through Telomerase Manipulation. Cell Rep *17*, 1184-1192.
Wang, F., Flanagan, J., Su, N., Wang, L.C., Bui, S., Nielson, A., Wu, X., Vo, H.T., Ma, X.J., and Luo, Y. (2012). RNAscope: a novel in situ RNA analysis platform for formalin-fixed, paraffin-embedded tissues. J Mol Diagn *14*, 22-29.

Zhang, Y., Chen, K., Sloan, S.A., Bennett, M.L., Scholze, A.R., O'Keeffe, S., Phatnani, H.P., Guarnieri, P., Caneda, C., Ruderisch, N., *et al.* (2014). An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. J Neurosci *34*, 11929-11947.

Figures titles and legends

Figure 1. Random pooling analysis identify stable genes specific for human SNc and VTA dopamine neurons.

(A) Hierarchical clustering of samples from the current study (18 male subjects, 74 DEGs).

(B) The identified 74 DEGs also separate SNc and VTA samples (three female subjects) from Nichterwitz et al., 2016.

(C) Venn-diagram of the DEGs between SNc and VTA dopamine neurons found independently for each dataset.

(D) Schematic of the algorithm used to identify stable genes, see STAR methods.

(E and F) Output of the algorithm showing the probability to find stable genes for the SNc (E) and VTA (F) as a function of sample size. For each subtype, genes with an asterisk (*) are common with overlapping DEGs (Venn-diagram in C).

(G) SNc and VTA stable genes segregate both dopamine neuron subtypes shown in a clustered heat map.

(H and I) RNAscope staining and quantification for the stable genes *SEZ6* (H, p=0.025) and *CDH13* (I, p=0.005) enriched in the SNc and VTA, respectively (n= 5 subjects, data represented as mean ± SEM, Paired t test).

Scale bars 30 μm (15 μm for insets). See also Figure S2.

Figure 2. Comparison of stable genes defining SNc and VTA dopamine neurons in mouse and man.

(A and B) Identification of SNc (A) and VTA (B) stable genes in the mouse dataset from La Manno et. al, 2016 using the random pooling algorithm. See STAR methods and Figure S3C.

(C) Venn-diagram representation showing the overlap between stable genes across mouse and man.

(D to G) PCAs for the mouse dataset from La Manno study when using: mouse stable genes (D), mouse stable list after exclusion of the three common genes (E), human stable genes (F) or human stable list after exclusion of the three common genes (G), The accuracy (Acc.) to distinguish those subpopulations is shown as a percentage value.

(H to K) Similar analysis as (D to G) but applied to human samples (subjects from our current study and Nichterwitz et al., 2016.

(L and M) The three common stable genes are sufficient to distinguish SNc and VTA samples in mouse (L) and human (M). See also Figure S3.

Figure 3. Identification and validation of cross-species SNc and VTA dopamine neuron predictive genes.

(A) Ranked mean accuracy for individual human stable genes, resolving the mouse dataset from La Manno et al., 2016, upon random exclusion together with any other two genes. Common stable genes between mouse and human are labelled in blue, while additional genes with predictive value in the mouse are labelled in green.
(B and C) *In situ* hybridization images of midbrain sections from the adult mouse (P56, Allen Brain Atlas) showing RNA expression pattern for *Th* (coronal in B, sagittal in C), the common markers *Serpine2*, *Zcch12* and *Cdh13* (in blue, B) and human stable genes with predictive value in the mouse (in green, C), in addition to the mouse VTA stable marker *Calb1* also useful for human. See Figure S4A. Scale bars: (B and C) 1mm for *Th* and 200 μm for remaining markers. See also Figure S4.

STAR METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCE TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- SUBJECT DETAILS
    - Ethics statement
- METHOD DETAILS
    - Tissue sectioning and laser capture
    - Mapping and gene expression quantification
    - Differentially expressed analyses
    - Random pooling algorithm
        - Random pooling applied to human samples
        - Random pooling applied to mouse single cells
    - Accuracy calculation
    - RNAscope staining
- STATISTICAL ANALYSIS
- DATA AVAILABILITY

CONTACT FOR REAGENT AND RESOURCE SHARING

As a lead contact, Eva Hedlund is responsible for all reagent and resource requests. Please contact Eva Hedlund at eva.hedlund@ki.se with requests and inquiries.

SUBJECT DETAILS

Ethics statement

We have ethical approval to work with human post-mortem samples (Tables S2 and S3) from the

regional ethical review board of Stockholm, Sweden (EPN Dnr 2012/111-31/1; 2012/2091-32). Fresh frozen tissue was obtained through the Netherlands Brain Bank (NBB) with the written informed consent from the donors or the next of kin. The work with human tissues was carried out according to the Code of Ethics of the World Medical Association (Declaration of Helsinki).

METHOD DETAILS

Tissue sectioning and laser capture

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

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

Mapping and gene expression quantification Samples were sequenced using an Illumina HiSeq2000 or HiSeq2500 platform (reads of 43 or 50 bp in length respectively). The uniquely mapped reads were obtained by mapping to the human reference genome hg38/GRCh38 using STAR with default settings. The reads per kilobase of transcript per million mapped reads (RPKM) were estimated using “rpkmforgenes” (Ramskold et al., 2009).

As part of the quality control we verified that samples had between 1.3~9.8 million reads and 7.9 ~ 12.3 kilo genes expressed with RPKM>1. Correlation between any two nearest samples was considered to be above 0.7. Additionally, we confirmed the expression on known midbrain dopamine neuron markers and the purity of each sample (see Figure S2H; Figure S3B).

Differentially expressed analyses

Differentially expressed genes were identified using the R package “DESeq2” (current version: 1.16.1) (Love et al., 2014) where the cutoff for significance was an adjusted P value of 0.05. Identified DEGs (from different analysis and summarized below) are shown in Tables S4 to S8.

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

Table S5: Human differentially expressed genes (current study): 74 DEGs calculated from 16 SNc and 14 VTA samples from 18 male individuals.

Table S6: Human differentially expressed genes (Nichterwitz et al., 2016): 100 DEGs calculated from 4 SNc and 3 VTA samples from 3 female subjects.

Table S7: Human stable genes: 33 DEGs calculated from 12 individuals (from the current study) using the random pooling algorithm.
Table S8: Mouse stable genes: 89 DEGs calculated from 73 SNc and VTA cells (La Manno et al., 2016) using the random pooling algorithm.

Random pooling algorithm

To counteract the variability among human subjects and identify the most reliable DEGs between SNc and VTA neurons across datasets we developed a random pooling algorithm (Figure 1D; Figure S3C). The stable genes output of this analysis is correlated with the sample size and give an unbiased estimation of the number of individuals required to consistently distinguish these closely related subpopulations. Importantly this computational tool can be used for the comparison of any other two groups (https://github.com/shanglicheng/RandomPooling, RRID:SCR_016607).

In details:

1) Define ¨N¨ and ¨M¨ as the number of samples in Group 1 and Group 2, respectively.
2) Define ¨n¨ and ¨m¨ as the number of randomly selected samples from Group1 and Group2, where $n \in \{3, 4, 5, \ldots, N\}$ and $m \in \{3, 4, 5, \ldots, M\}$.
3) Pool ¨n¨ and ¨m¨ samples (temporary considered a ¨new data set¨) and calculate DEGs.
4) Repeat steps 2) and 3) for ¨j¨ times (set to 1000 times in this study).
5) For every round, save the full list of DEGs, compute and rank their frequency.
6) Set a threshold (30% ratio in this study) to consider DEGs with higher frequency as stable genes.
Reliable genes appear when frequencies are above: Total times of random pooling x ratio (300 in this study). A stringent, but fair ratio can be defined by comparing the percentage of identified stable genes overlapping with the top 10%, 20%, 30% DEGs.

Random pooling applied to human samples

To reliable identify DEGs between human SNc and VTA samples, while minimizing subject variability, we selected 12 individuals (66% of the dataset, 12 out of 18 individuals) where both neuronal groups were available and sequenced. Hence, the number of randomly selected samples (¨n¨ and ¨m¨ from ¨i¨ individuals) was from three to 11 and the algorithm repeated 1000 times (Figure 1D).

Random pooling applied to mouse single cells

For this adult mouse dataset (La Mano et. al, 2016) we defined the groups SNc (N=73 cells) and VTA (M=170 cells comprising VTA1, VTA2, VTA3 and VTA4). To compensate the unbalance in cell number and adjust dataset representation compared to the human analysis (66%), we first

randomly collected a subset of 73 VTA cells, pairing both SNc and VTA. Similarly, the number of randomly selected samples (¨n¨ and ¨m¨ from ¨i¨ cells) was from 20 to 70 and the algorithm repeated again 1000 times (Figure S3C).

Accuracy calculation

To measure the power of stable genes in resolving SNc and VTA samples we calculated the accuracy. This is a quantitative parameter associated with correct group prediction using hierarchical clustering based on the first two dimensions (PC1 and PC2) of the PCA visualizing a given dataset. As we have two predicted groups (SNc and VTA), the minimum value for the accuracy is 50%. In details:

1) Define ¨N¨ and ¨M¨ as the total number of samples in Group 1 and Group 2, respectively.
2) Define ¨n¨ and ¨m¨ as the number of correctly predicted samples for Group1 and Group2, respectively.
3) Accuracy, expressed as percent, is calculated according to the formula:

Accuracy (%)= ((¨n¨+¨m¨)/(¨N¨+¨M¨)) * 100

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

a) Mouse stable genes applied to mouse (Figure 2D) or human (Figure 2J) samples.
b) Mouse stable genes excluding (*Zcchc12*, *Cdh13*, *Serpine2*) applied to mouse (Figure 2E) or human (Figure 2K) samples.
c) Human stable genes applied to mouse (Figure 2F) or human (Figure 2H) samples.
d) Human stable genes excluding (*ZCCHC12*, *CDH13*, *SERPINE2*) applied to mouse (Figure 2G) or human (Figure 2I) samples.
e) Mouse stable genes excluding three random markers applied to human samples (Figure S4A).
f) Human stable genes excluding three random markers applied to mouse samples (Figure 3A).

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

RNAscope staining of human tissues

RNAscope (Wang et al., 2012) was used to verify the expression of one SNc marker (*SEZ6*) and one VTA-preferential gene (*CDH13*) based on the sequencing data. In brief midbrain sections of human fresh frozen tissue (Table S3) were quickly thawed and fixed with fresh PFA (4% in

PBS) for 1 hour at 4'C. The RNAscope 2.5 HD Assay - RED Kit (Cat. 322360) was used using manufacturer recommendations.

To evaluate the procedure in the midbrain tissue (Figure 2N), we first tested a negative control probe against a bacterial gene (Cat. 310043, *dapB*-C1) and a positive control probe against tyrosine hydroxylase (Cat. 441651, *TH*-C1). Once we set up the assay, midbrain sections were stained with *SEZ6* (Cat. 411351-C1) or *CDH13* probes (Cat. 470011-C1). Slides were counterstained with fresh 50% Gill No.1 Hematoxylin (Cat. GSH132-1L, Sigma-Aldrich) for 2 minutes, washed in water and dried for 15min at 60°C before mounting with Pertex (Cat. 00811, Histolab).

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

STATISTICAL ANALYSIS

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

DATA AVAILABILITY

RNA sequencing data generated in this study have been submitted to the Gene Expression Omnibus (GEO) of the National Center for Biotechnology Information under the accession number GSE114918. Human samples re-analyzed from the Nichterwitz study (Nichterwitz et al., 2016) are under accession number GSE76514 and mouse adult single cells (La Manno et al., 2016) can be accessed at GSE76381. Previous human microarray studies used to verify SNc stable gene expression were accessed from the European repository ArrayExpress (E-MEXP-1416) (Cantuti-Castelvetri et al., 2007) or raw data requested to Dr. Kai C. Sonntag (Simunovic et al., 2009).

Supplemental information titles and legends

Figure S1. Common differentially expressed genes across species. Related to tables S1 and S4.

(A) Venn diagram showing overlapping DEGs for the SNc and VTA identified in previous rat studies (Greene et al., 2005; Grimm et al., 2004).

(B) Common mouse DEGs are highlighted for the SNc and VTA from three different reports (Bifsha et al., 2014; Chung et al., 2005; La Manno et al., 2016).

(C) Overlapping DEGs between rat, mouse and human (Nichterwitz et al., 2016).

Figure S2. Gene signature of human adult midbrain dopamine neurons using LCM-seq. Related to Figure 1 and Tables S2, S5 and S7.

(A to G) Nissl staining of a representative midbrain section for the isolation of SNc (B,C,D) or VTA (E,F,G) neurons before and after laser capture.

(H) Quality of the samples profiled in the current study was confirmed by the expression of known midbrain dopamine neuron markers and neurofilament (*NEFH*), and the lack of astrocyte, microglia or oligodendroyte precursor contamination (Zhang et al., 2014).

(I and J) Samples from the Nichterwitz's study (Nichterwitz et al., 2016) were re-analyzed with DESeq (version 1.16.1) which identified 100-DEGs used to represent those samples (I) or our current dataset (J).

(K and L) Frequency histograms (ranked by p-value), output of the random pooling algorithm, extend the analysis of identified SNc and VTA stable genes in human. Genes with the highest frequency (at the top, in red) represent SNc (K) and VTA (L) stable genes. Among those: *GSG1L* (Nichterwitz et al., 2016; Poulin et al., 2014), *SLIT1* (Nichterwitz et al., 2016), *SLC17A6* (Chung et al., 2005; Grimm et al., 2004; Nichterwitz et al., 2016) and *EN2* (Nichterwitz et al., 2016; Poulin et al., 2014) were previously reported as differentially expressed. Many other genes reported as differentially expressed were also found in the histograms, but at a lower frequency (not permissive to consider as stable).

(M) SNc stable genes are expressed in control human SNc dopamine neurons isolated by laser capture microdissection from previous microarray studies. [#]In the study reported by Simunovic et al. (Simunovic et al., 2009), the genes *GSG1L* and *SEZ6* were not included in the array.

(N) RNAscope staining of human melanized midbrain sections using a negative control probe against a bacterial gene (*dapB*) or a positive control probe against tyrosine hydroxylase (*TH*). Scale bars: 500 μm in (A), 200 μm in (B, E), 50 μm in (C,D,F,G) and 1mm in (N) with 50 μm in insets.

Figure S3. Identification of stable genes in the mouse SNc and VTA dopamine neurons. Related to Figure 2 and Table S8.

(A) Mouse dataset used to identify stable genes in juvenile/adult (postnatal day 19 to 56) midbrain dopamine neurons. The graph shows the number of single cells sequenced for each dopaminergic subtype.

(B) Heat map to confirm the expression of known midbrain dopamine neuron markers and the absence of markers for other neural cells as astrocytes, microglia or oligodendrocytes.

(C) Schematic of the random pooling algorithm used to identify mouse stable genes from the single cell data.

(D) Identified mouse stable genes separate SNc and VTA samples as visualized in the unsupervised clustering.

(E and F) Frequency histograms (ranked by p-value) output of the random pooling algorithm. Identified mouse SNc (E) and VTA (F) stable genes are represented at the top (in red). Some of those stable markers have been previously identified as differentially expressed (Bifsha et al., 2014; Chung et al., 2005; Greene et al., 2005; Grimm et al., 2004; La Manno et al., 2016; Nichterwitz et al., 2016; Poulin et al., 2014). Other reported DEGs are found in the histograms but not considered stable due to their low frequency.

Figure S4. Identification and validation of mouse SNc and VTA dopamine neuron predictive genes. Related to Figure 3 and Table S8.

(A) Ranked mean accuracy for individual mouse stable genes upon random exclusion together with any other two genes to resolve human samples (current study plus Nichterwitz et al. samples) (Nichterwitz et al., 2016). Common stable genes between mouse and human are labelled in blue while additional genes with some predictive value in the human (led by *Calb1*) are labelled in green.

(B and C) *In situ* hybridization images of coronal midbrain sections of the adult mouse (P56, Allen Brain Atlas) showing RNA expression pattern for *Th*, the common genes *Serpine2*, *Zcch12* and *Cdh13* (in blue) and mouse stable genes with some predictive value for human (in green). The expression pattern of some selected stable genes with restricted predictive value to the mouse was also verified by *in situ* hybridization (black labels).

Scale bars: (B and C) 1mm for *Th* and 200 μm for remaining markers.

Table S1. Summary of studies addressing the identity of midbrain dopamine neurons. Related to Figure S1.

Table S2. Clinical information of subjects used for LCM-seq. Related to Figures 1 and S2.

Table S3. Clinical information of subjects used for RNAscope. Related to Figures 1 and S2.

Tables S4 to S8. List of DEGs and Stable Genes (combined in 1 excel file with 5 sheets)

Table S4. List of differentially expressed genes from previous transcriptomic studies. Related to Figure S1.

Table S5. List of differentially expressed genes from the current dataset (18 male subjects, 74 DEGs). Related to Figures 1 and S2.

Table S6. List of differentially expressed genes from the Nichterwitz's dataset (Nichterwitz et al., 2016) re-analized here (3 female subjects, 100 DEGs). Related to Figures 1 and S2.

Table S7. Full list of human stable genes/DEGs ranked by frequency identified with the random pooling algorithm. See also STAR METHODS section. Related to Figures 1, 2, 3 and Figure S2.

Table S8. Full list of mouse stable genes/DEGs ranked by frequency identified with the random pooling algorithm. See also STAR METHODS section. Related to Figure 2 and Figures S3 and S4.

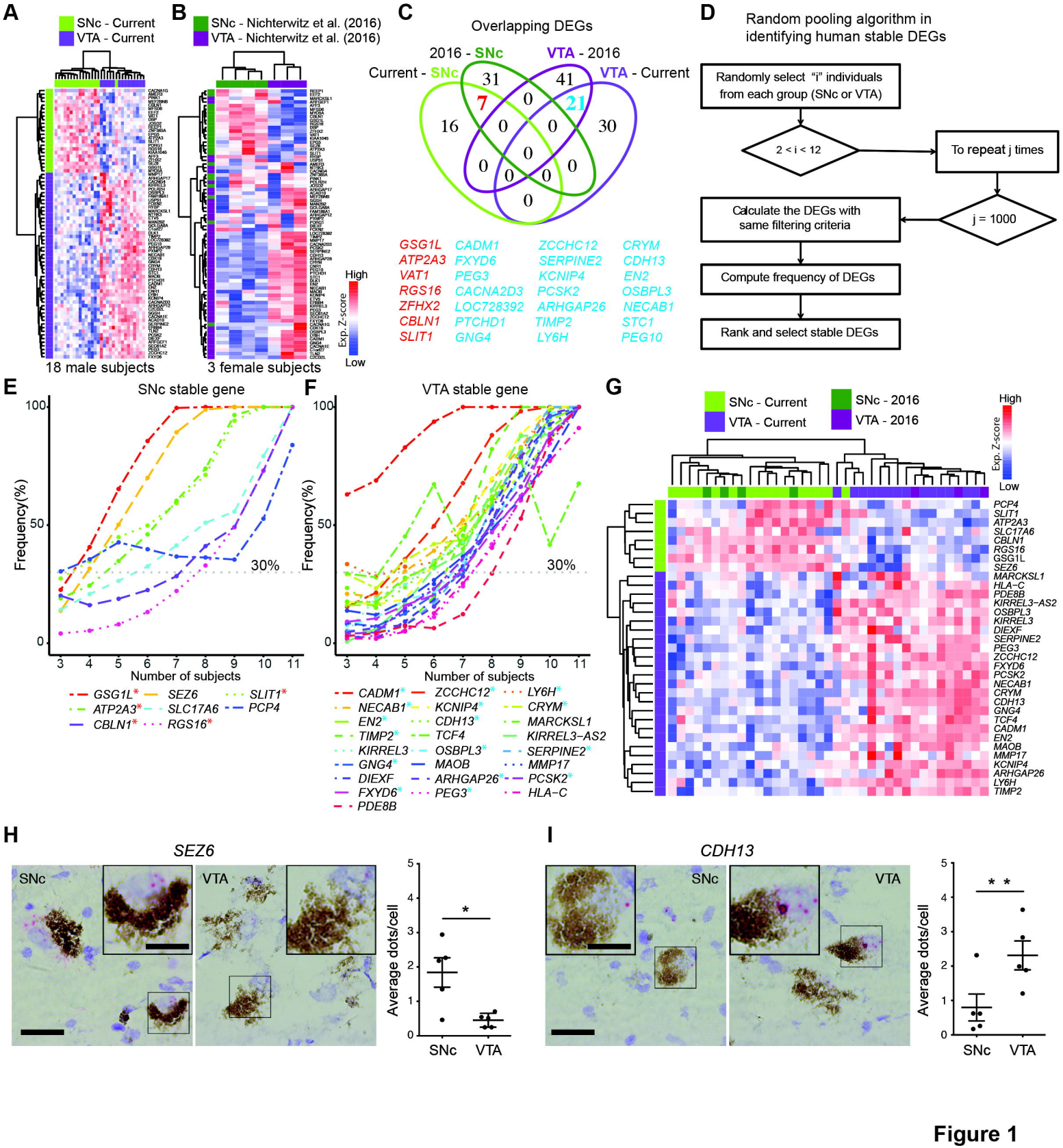
A
SNc - Current
VTA - Current
18 male subjects
B
SNc - Nichterwitz et al. (2016)
VTA - Nichterwitz et al. (2016)
3 female subjects
High
Exp. Z-score
Low
C
Overlapping DEGs
2016 - SNc
VTA - 2016
Current - SNc
VTA - Current
31
41
7
21
16
30
0
GSG1L ATP2A3 VAT1 RGS16 ZFHX2 CBLN1 SLIT1
CADM1 FXYD6 PEG3 CACNA2D3 LOC728392 PTCHD1 GNG4
ZCCHC12 SERPINE2 KCNIP4 PCSK2 ARHGAP26 TIMP2 LY6H
CRYM CDH13 EN2 OSBPL3 NECAB1 STC1 PEG10
D
Random pooling algorithm in identifying human stable DEGs
Randomly select "i" individuals from each group (SNc or VTA)
2 < i < 12
To repeat j times
j = 1000
Calculate the DEGs with same filtering criteria
Compute frequency of DEGs
Rank and select stable DEGs
E
SNc stable gene
Frequency(%)
100
50
0
30%
3 4 5 6 7 8 9 10 11
Number of subjects
GSG1L* SEZ6 SLIT1*
ATP2A3* SLC17A6 PCP4
CBLN1* RGS16*
F
VTA stable gene
Frequency(%)
100
50
0
30%
3 4 5 6 7 8 9 10 11
Number of subjects
CADM1* ZCCHC12* LY6H*
NECAB1* KCNIP4* CRYM*
EN2* CDH13* MARCKSL1
TIMP2* TCF4 KIRREL3-AS2
KIRREL3 OSBPL3* SERPINE2*
GNG4* MAOB MMP17
DIEXF ARHGAP26* PCSK2*
FXYD6* PEG3* HLA-C
PDE8B
G
SNc - Current
VTA - Current
SNc - 2016
VTA - 2016
High
Exp. Z-score
Low
PCP4
SLIT1
ATP2A3
SLC17A6
CBLN1
RGS16
GSG1L
SEZ6
MARCKSL1
HLA-C
PDE8B
KIRREL3-AS2
OSBPL3
KIRREL3
DIEXF
SERPINE2
PEG3
ZCCHC12
FXYD6
PCSK2
NECAB1
CRYM
CDH13
GNG4
TCF4
CADM1
EN2
MAOB
MMP17
KCNIP4
ARHGAP26
LY6H
TIMP2
H
SEZ6
SNc
VTA
Average dots/cell
5
4
3
2
1
0
*
SNc VTA
I
CDH13
SNc
VTA
Average dots/cell
5
4
3
2
1
0
* *
SNc VTA

Figure 1

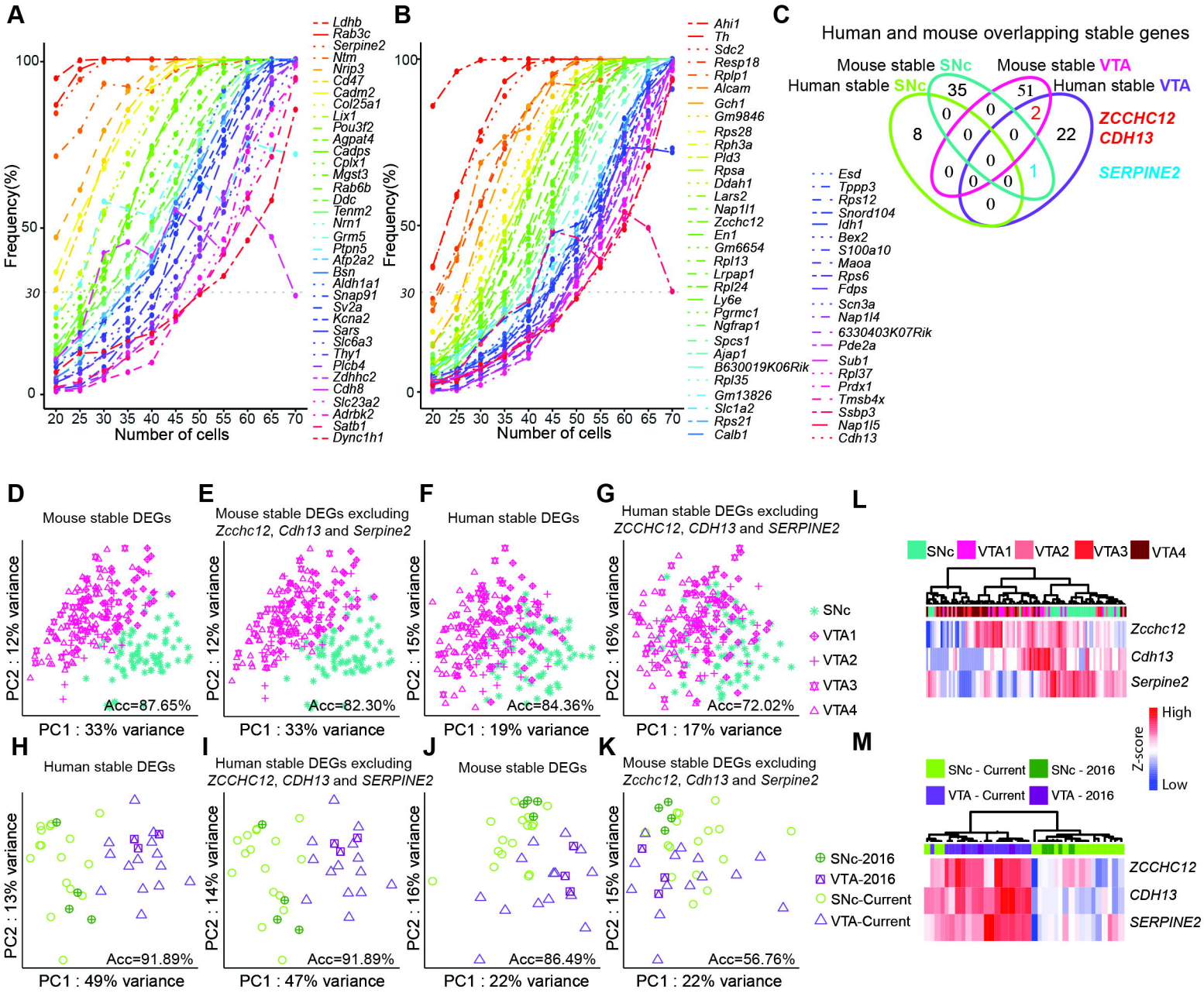
A
Frequency(%)
Number of cells
Ldhb
Rab3c
Serpine2
Ntm
Nrip3
Cd47
Cadm2
Col25a1
Lix1
Pou3f2
Agpat4
Cadps
Cplx1
Mgst3
Rab6b
Ddc
Tenm2
Nrn1
Grm5
Ptpn5
Atp2a2
Bsn
Aldh1a1
Snap91
Sv2a
Kcna2
Sars
Slc6a3
Thy1
Plcb4
Zdhhc2
Cdh8
Slc23a2
Adrbk2
Satb1
Dync1h1
B
Ahi1
Th
Sdc2
Resp18
Rplp1
Alcam
Gch1
Gm9846
Rps28
Rph3a
Pld3
Rpsa
Ddah1
Lars2
Nap1l1
Zcchc12
En1
Gm6654
Rpl13
Lrpap1
Rpl24
Ly6e
Pgrmc1
Ngfrap1
Spcs1
Ajap1
B630019K06Rik
Rpl35
Gm13826
Slc1a2
Rps21
Calb1
C
Human and mouse overlapping stable genes
Mouse stable SNc
Mouse stable VTA
Human stable SNc
Human stable VTA
ZCCHC12
CDH13
SERPINE2
Esd
Tppp3
Rps12
Snord104
Idh1
Bex2
S100a10
Maoa
Rps6
Fdps
Scn3a
Nap1l4
6330403K07Rik
Pde2a
Sub1
Rpl37
Prdx1
Tmsb4x
Ssbp3
Nap1l5
Cdh13
D
Mouse stable DEGs
PC2 : 12% variance
PC1 : 33% variance
Acc=87.65%
E
Mouse stable DEGs excluding Zcchc12, Cdh13 and Serpine2
PC2 : 12% variance
PC1 : 33% variance
Acc=82.30%
F
Human stable DEGs
PC2 : 15% variance
PC1 : 19% variance
Acc=84.36%
G
Human stable DEGs excluding ZCCHC12, CDH13 and SERPINE2
PC2 : 16% variance
PC1 : 17% variance
Acc=72.02%
SNc
VTA1
VTA2
VTA3
VTA4
H
Human stable DEGs
PC2 : 13% variance
PC1 : 49% variance
Acc=91.89%
I
Human stable DEGs excluding ZCCHC12, CDH13 and SERPINE2
PC2 : 14% variance
PC1 : 47% variance
Acc=91.89%
J
Mouse stable DEGs
PC2 : 16% variance
PC1 : 22% variance
Acc=86.49%
K
Mouse stable DEGs excluding Zcchc12, Cdh13 and Serpine2
PC2 : 15% variance
PC1 : 22% variance
Acc=56.76%
SNc-2016
VTA-2016
SNc-Current
VTA-Current
L
SNc
VTA1
VTA2
VTA3
VTA4
Zcchc12
Cdh13
Serpine2
High
Z-score
Low
M
SNc - Current
SNc - 2016
VTA - Current
VTA - 2016
ZCCHC12
CDH13
SERPINE2

Figure 2

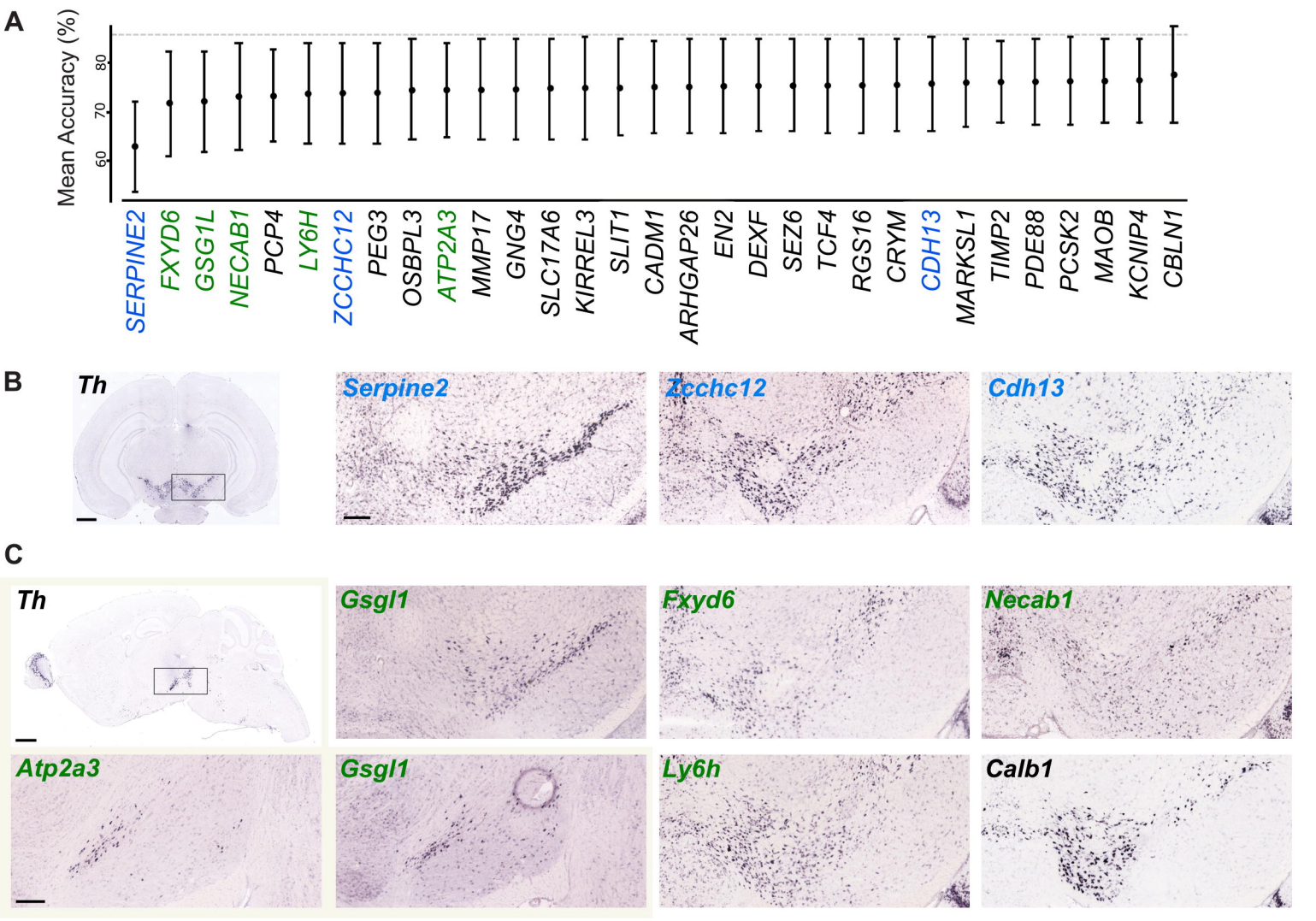
A
Mean Accuracy (%)
60
70
80
SERPINE2
FXYD6
GSG1L
NECAB1
PCP4
LY6H
ZCCHC12
PEG3
OSBPL3
ATP2A3
MMP17
GNG4
SLC17A6
KIRREL3
SLIT1
CADM1
ARHGAP26
EN2
DEXF
SEZ6
TCF4
RGS16
CRYM
CDH13
MARKSL1
TIMP2
PDE88
PCSK2
MAOB
KCNIP4
CBLN1
B
Th
Serpine2
Zcchc12
Cdh13
C
Th
Gsgl1
Fxyd6
Necab1
Atp2a3
Gsgl1
Ly6h
Calb1

Figure 3