

1 **Supplementary Materials**

2

3 **Molecular heterogeneity of the non-human primate cochlea**

4

5 Xin Chen, Yuwei Che, Jieyu Qi, Ming Cen, Shan Gao, Bin Zhu, Yiheng Ao, Xiangyu Ma, Cheng
6 Cheng, Rongrong Guo, Xinlin Wang, Tian Shen, Jiheng Wu, Fanliang Kong, Wei Si, Yongchang
7 Chen, Tao Tan, Ling Lu, Renjie Chai

8

9 This file includes:

10

11 Supplementary Figures 1-15

12 Supplementary Data 1-2

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

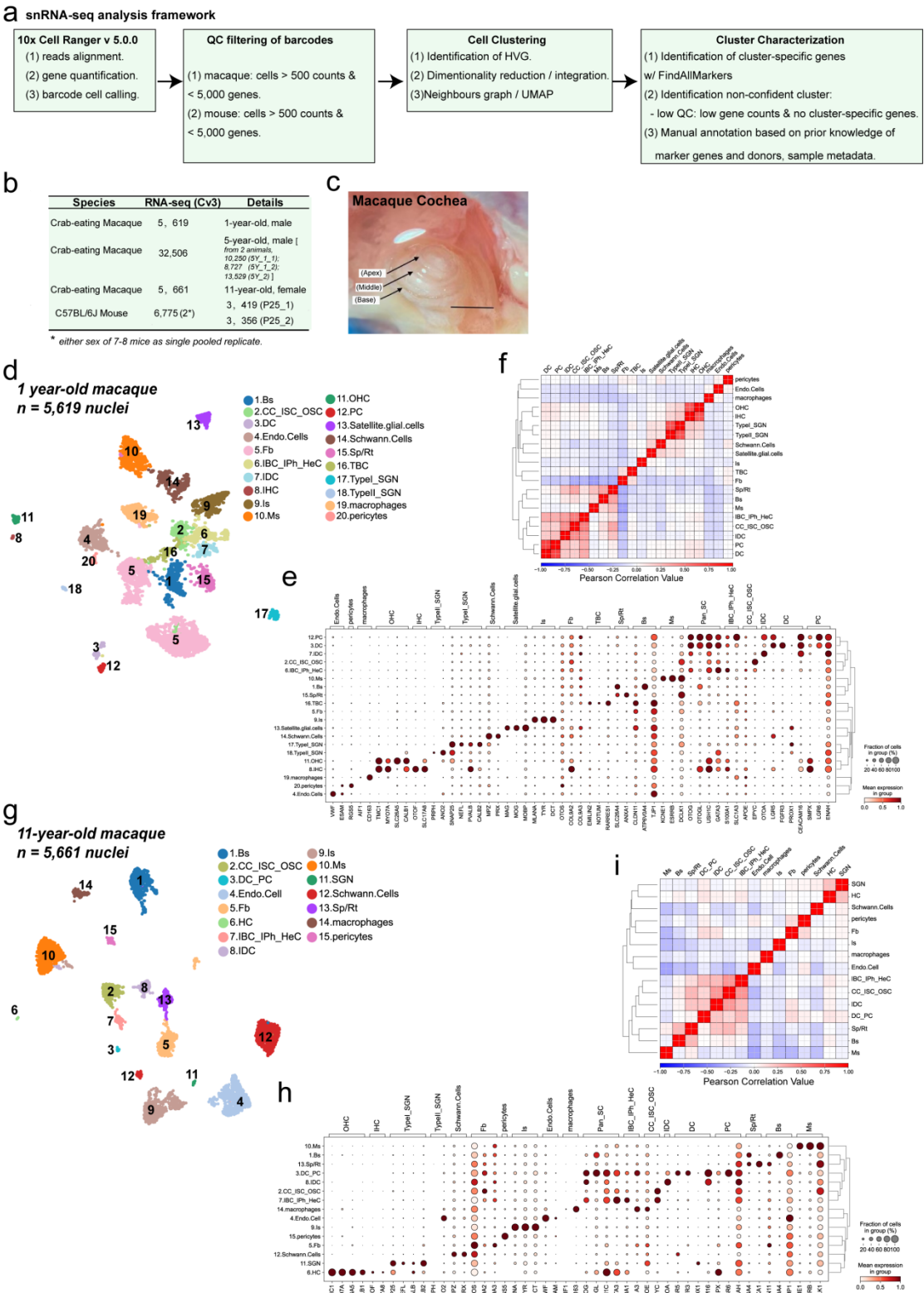
39

40

41

42

43 **Supplementary Figures**



44

45

46

47

Supplementary Figure 1
Sample overview and quality control metrics of 1-year-old and 11-year-old

macaque cochleae related to Figure 1.

a Schematic of the snRNA-seq (single-nucleus RNA sequencing) workflow for nuclei collection and data analysis.

b A table summarizing sample information, including species, collected nuclei after quality control, biological replicates, and sex.

c Bright-field image of the anatomical structure of a 1-year-old macaque cochlea, with insets indicating the approximate regions of the apical, middle, and basal turns. Scale bar = 300 μ m.

d Uniform Manifold Approximation and Projection (UMAP) plot of the 1-year-old macaque cochlea identified 20 distinct cell types. A total of 5,619 high-quality nuclei were captured prior to subsequent integration. These cell types include basal stria cells (Bs), root cells (Rt), intermediate stria cells (Is), spindle cells (Sp), marginal stria cells (Ms), endothelial cells (Endo.cells), pericytes, smooth muscle cells, pillar cells (PC), Deiters's cells (DC), inner border cells/inner phalangeal cells/Hensen's cells (IBC, IPh, HeC), interdental cells (IDC), Claudius cells/inner sulcus cells/outer sulcus cells (CC, ISC, OSC), inner hair cells (IHC), outer hair cells (OHC), type I spiral ganglion neurons (TypeI SGN), type II spiral ganglion neurons (TypeII SGN), satellite glial cells, Schwann cells, osteocytes (OB), Reissner's membrane (RM), tympanic border cells (TBC), and fibrocytes (Fb).

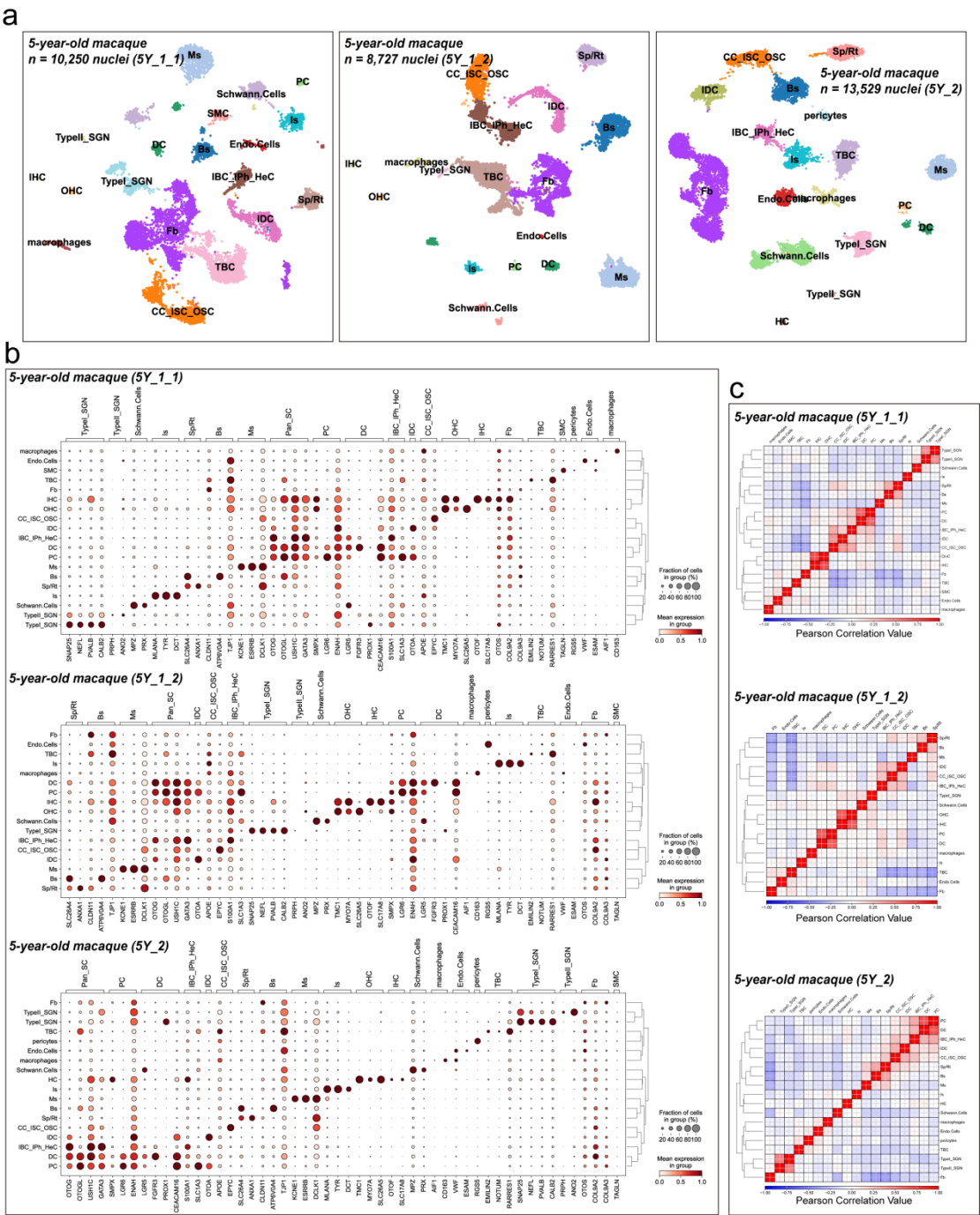
e A dot plot displaying the variance-scaled, log-transformed expression of marker genes (*x-axis*) identified in mouse species as a reference for mapping 1-year-old macaque cochlear cell types (*y-axis*). The main markers for manual annotation are from *Jean et al. PNAS, 2023*.

f Pearson correlation heatmap showing the similarity of transcriptomic expression patterns within cell types in the macaque cochlea (1-year-old).

g Same as (d), but for the 11-year-old dataset.

h Same as (e), but for the 11-year-old dataset.

i Same as (f), but for the 11-year-old dataset.

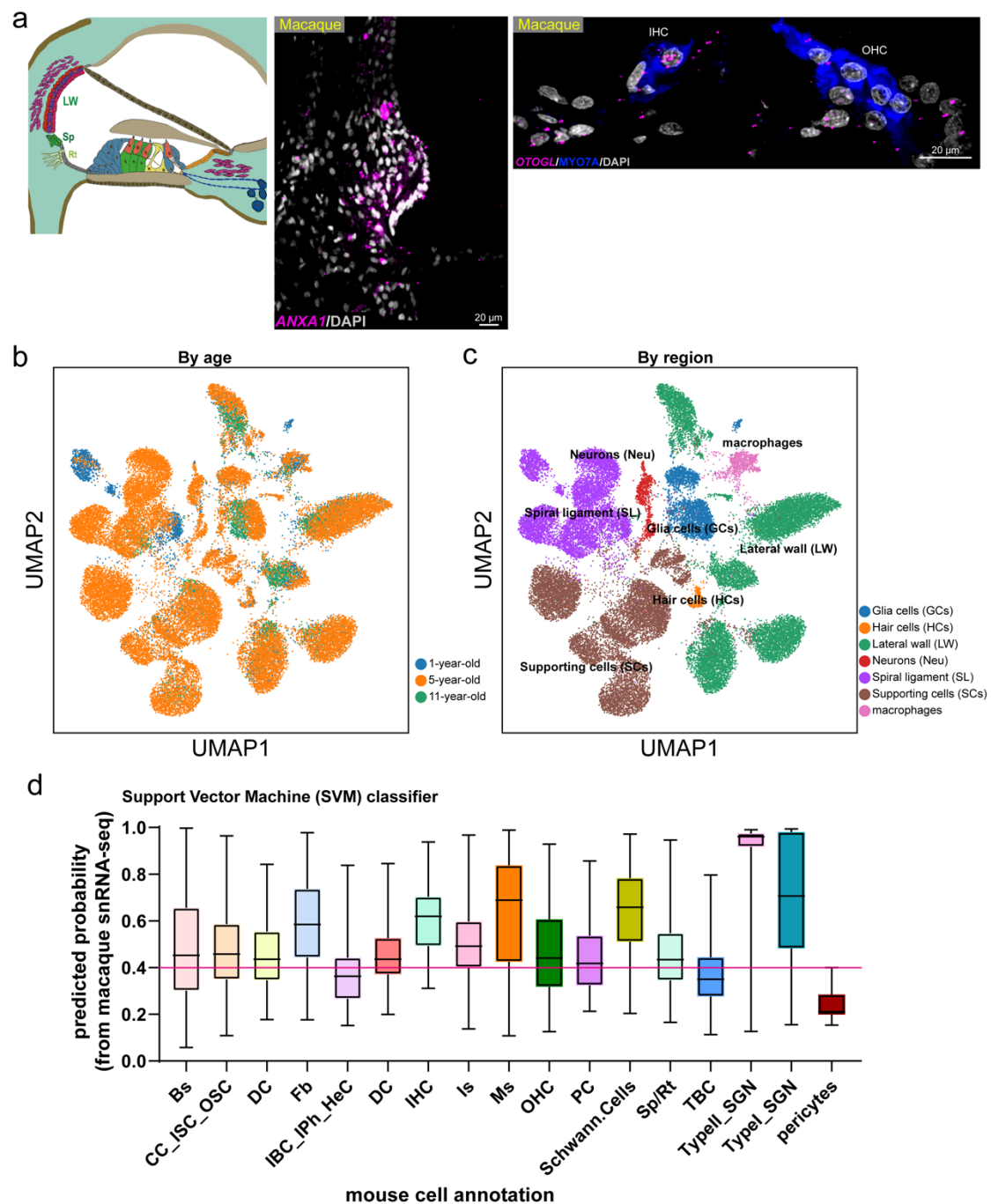


Supplementary Figure 2
Sample overview and quality control metrics of 5-year-old macaque cochlea related to Figure 1.

a Uniform Manifold Approximation and Projection (UMAP) plot of the 5-year-old macaque cochlear single-cell transcriptomes, which was identified more than 20 distinct cell types. A total of 32,506 high-quality nuclei from three individual cochleae were captured prior to subsequential integration.

b A dot plot displaying the variance-scaled, log-transformed expression of marker genes (x -

axis) identified in mouse species used for manual cell-type annotation for the 5-year-old macaque cochlea (*y-axis*). The marker gene list is mainly from *Jean et al. PNAS, 2023*.
c Pearson correlation heatmap showing the similarity of transcriptomic expression patterns within cell types in the macaque cochlea (5-year-old).



Supplementary Figure 3

Single-nucleus analysis of macaque cochlea, related to Figure 1.

a RNASCOPE assay validates *ANXA1* and *OTOGL* RNA molecules in 5-year-old macaque cochlea. The cellular components in the whole cochlea organ illustrated in a panel adopted and modified by Jean *et al* PNAS 2023.

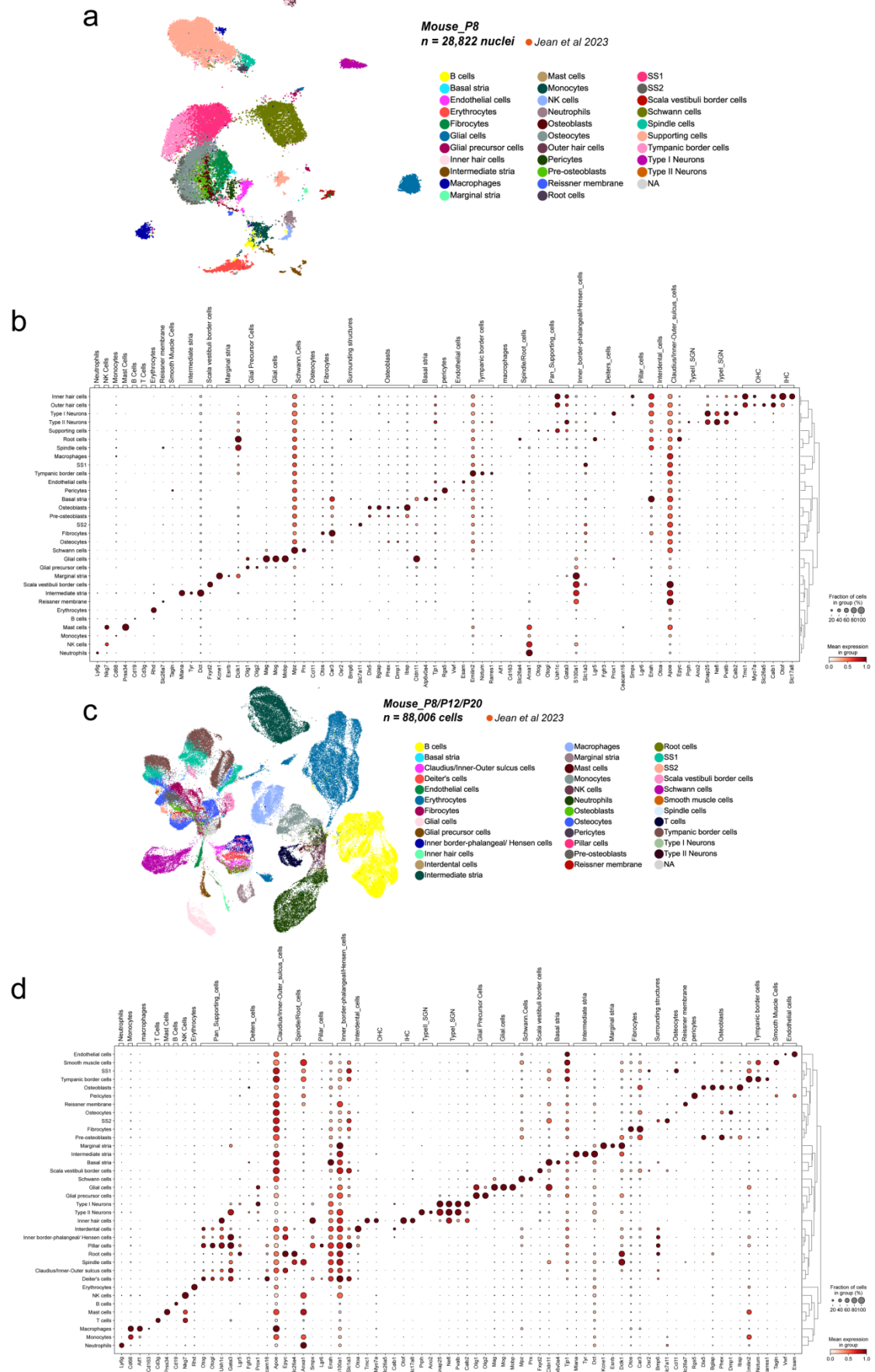
b Uniform Manifold Approximation and Projection (UMAP) plot of the integrative macaque cochlear single-nucleus transcriptomes, which are from 1-, 5-, and 11-year-old macaque cochleae. We used scVI (*single-cell variation inference*) for data integration and batch correction.

c UMAP plot displaying the main cochlear regions, including the cochlear epithelium (hair cells and supporting cells), spiral ganglion (neurons and glial cells), lateral walls, spiral ligament and macrophages.

d Boxplot showing the predicted probabilities (*y-axis*) of macaque cell types that were label-transferred onto manually postnatal days (P)25 mouse cell types (*x-axis*) with a Support Vector Machine (SVM) model as a classifier trained on the macaque cells. The box extends from the lower to upper quartile values of the data, with a line indicating a median value of 0.4.

203

204



Supplementary Figure 4

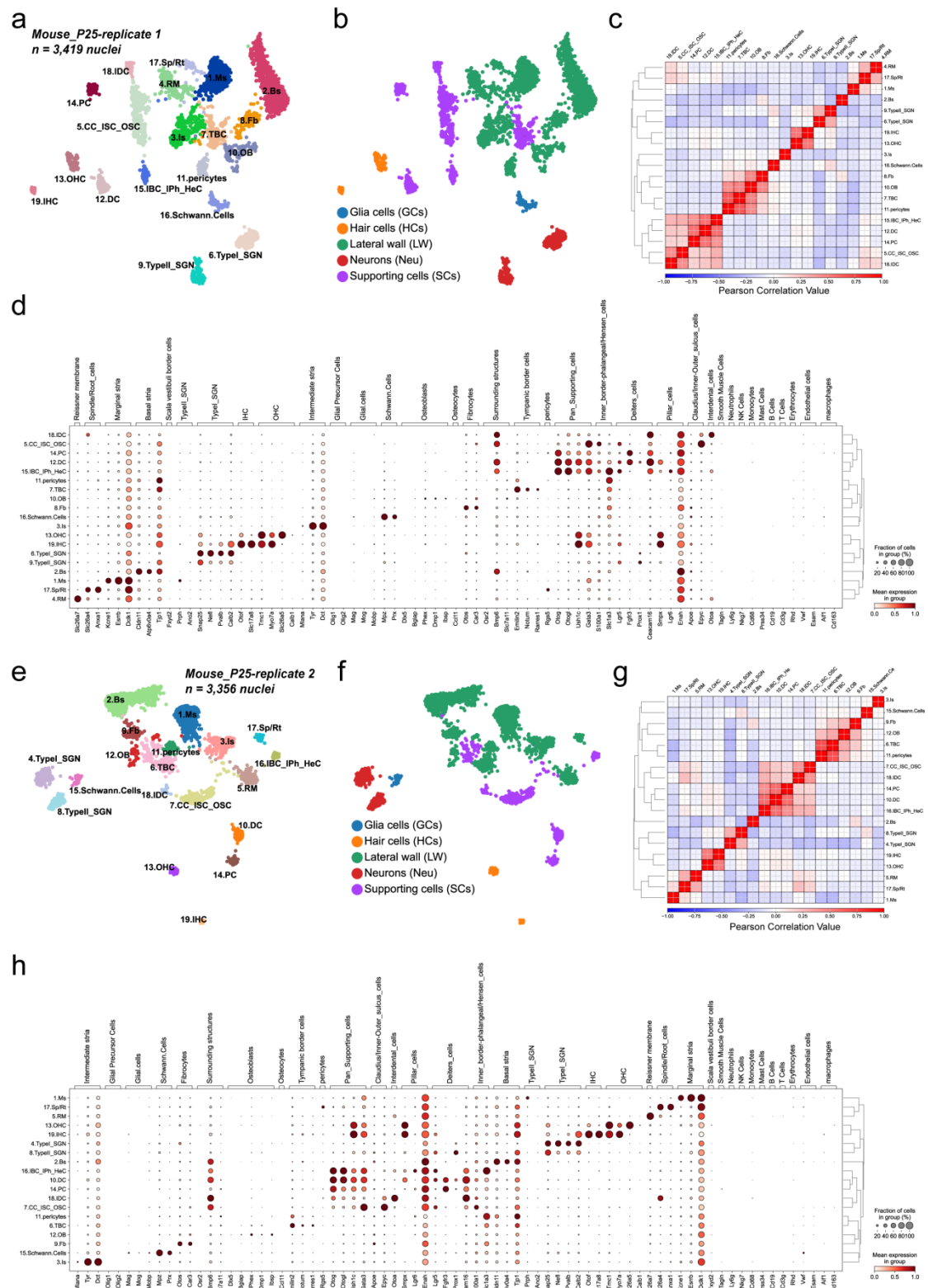
Single-cell or single-nucleus transcriptomic analysis of mouse cochlear cell types from PNAS, related to Figure 1.

a UMAP (Uniform Manifold Approximation and Projection) plot representing cell types of postnatal days (P)8 mouse cochleae from *Jean et al. PNAS, 2023*.

b Dot plot depicting the variation-scaled, log-transformed expression levels of marker genes (*x-axis*) in each cell type (*y-axis*) from P8 mouse cochleae.

c UMAP plot, similar to (**a**), but for pooled P8, P12, and P20 mouse cochleae. The capture method employed single-cell dissection and resuspension, and cell-type annotation was performed by *Jean et al. PNAS, 2023*.

d Same as (**b**), but for P8, P12, and P20 single-cell mouse cochleae.



Supplementary Figure 5
Single-nucleus transcriptomic profiling of the mouse cochlea, related to Figure 1.

a, b UMAP (Uniform Manifold Approximation and Projection) plots showing cell-type annotation (a) and main regions (b) of the postnatal days (P)25 mouse cochlea from sample replicate 1.

c Pearson correlation heatmap illustrating the similarity of transcriptomic expression patterns within cell types in the mouse cochlea of sample replicate 1.

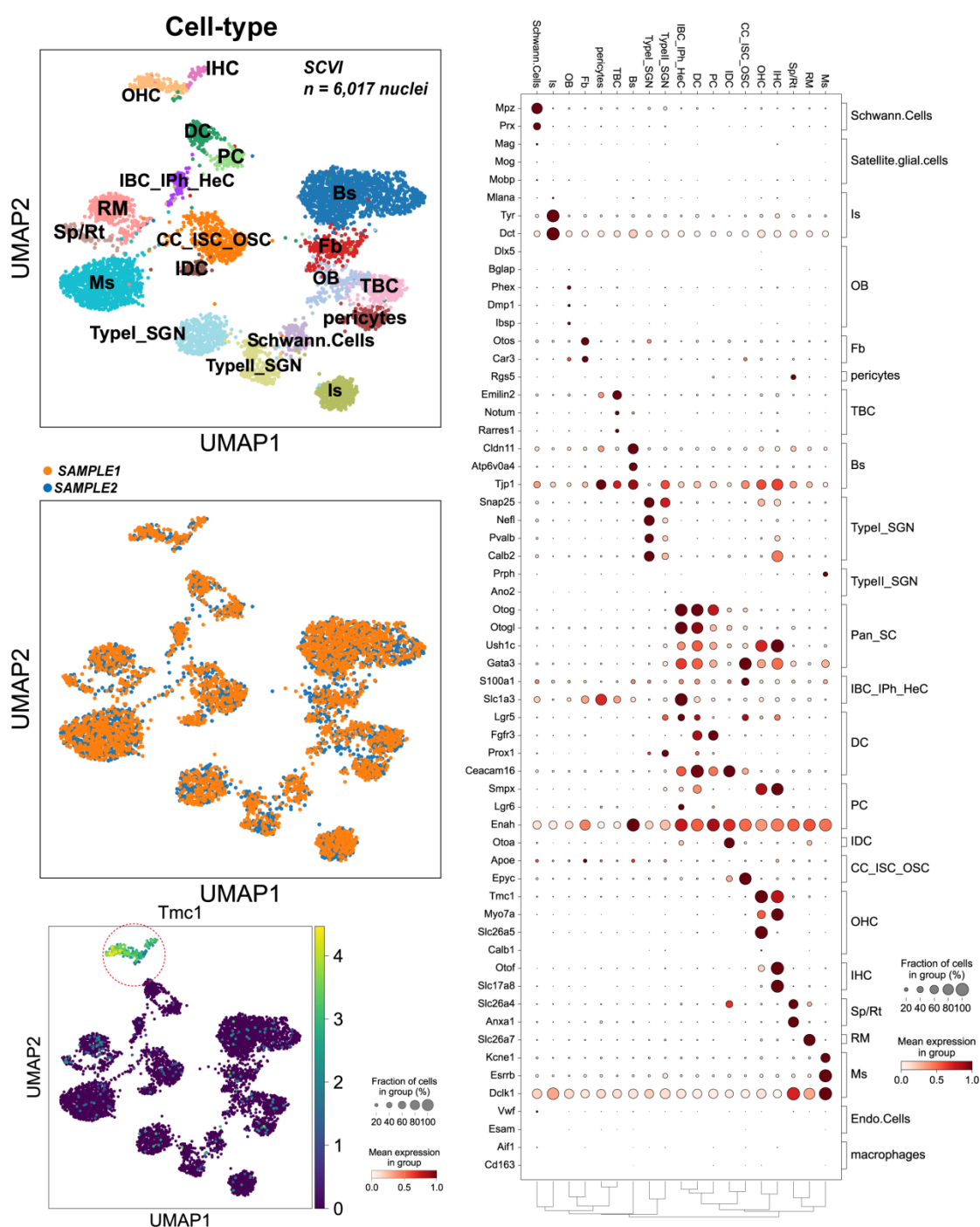
d Dot plot depicting the log-transformed expression levels of marker genes (*x-axis*) in each cell type (*y-axis*) from P25 mouse cochlea of sample replicate 1.

e, f Same as (**a, b**), but for sample replicate 2.

g Same as (**c**), but for sample replicate 2.

h Same as (**d**), but for sample replicate 2.

MOUSE

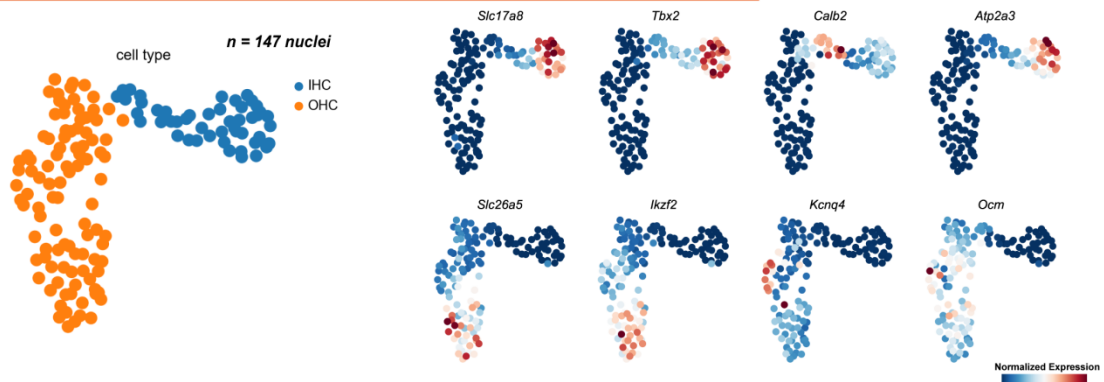


Supplementary Figure 6
Single-nuclei transcriptomic profiling of mouse cochlear cell types by
scVI integration, related to Figure 2.

Left: UMAP (Uniform Manifold Approximation and Projection) plots representation of cell types (up left), sample replicates (middle left), and highlighted hair cells marked by *Tmc1* (*transmembrane channel-like protein 1*) expression (bottom left) from postnatal days (P)25 mouse cochleae. Data combination of two sample replicates was imputed by scVI (*single-cell variation inference*) learning.

Right: dot plot depicting the variation-scaled, log-transformed expression levels of marker genes (*y-axis*) in each cell type (*x-axis*) from P25 mouse cochleae.

MOUSE

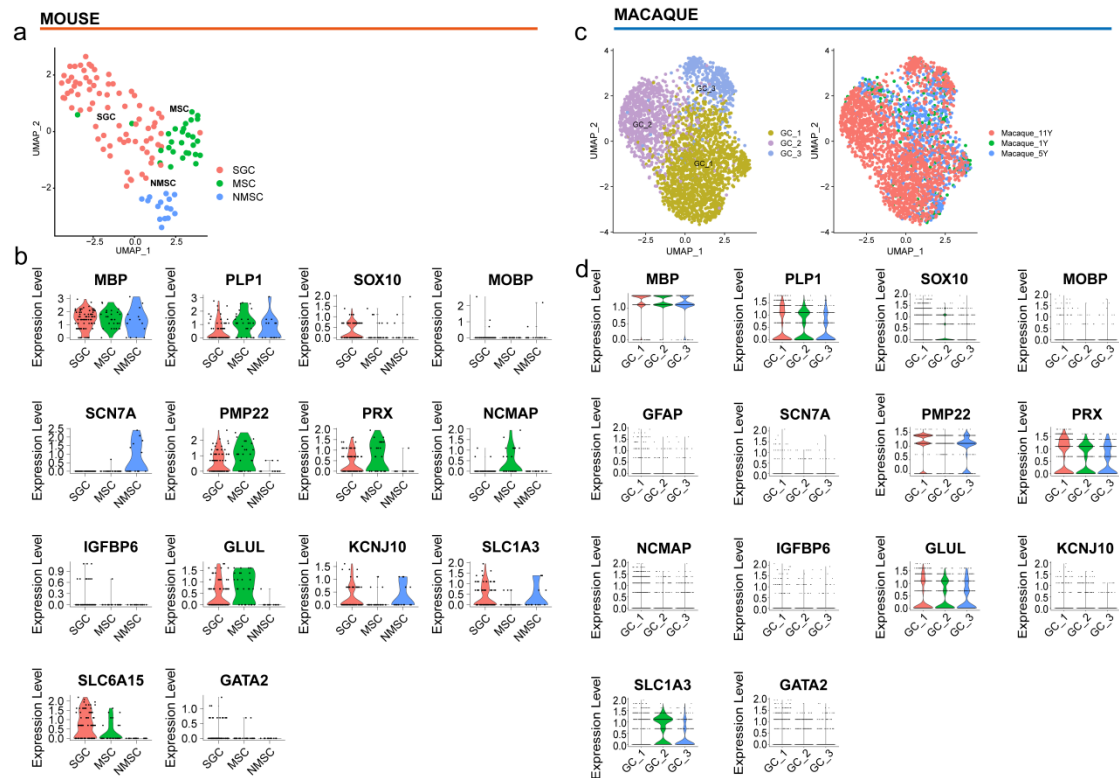


Supplementary Figure 7

Partitioning and re-clustering of mouse hair cell types, related to Figure 2.

Left: UMAP (Uniform Manifold Approximation and Projection) plot showing the integrated inner and outer hair cells of mouse hair cells.

Right: the canonical inner hair cell markers (*Slc17a8*, *Tbx2*, *Calb2*, *Atp2a3*) and outer hair cell markers (*Slc26a5*, *Ikzf2*, *Kcnq4*, *Ocm*) were distinct expressed across the two populations. The gene expression levels were calculated with log-normalization. Full-name of these genes is: *Slc17a8* (vesicular glutamate transporter 3), *Tbx2* (T-box transcription factor *TBX2*), *Calb2* (calretinin), *Atp2a3* (sarcolemmal/endoplasmic reticulum calcium ATPase 3), *Slc26a5* (solute carrier family 26 member 5, *Prestin*), *Ikzf2* (Ikaros family zinc finger protein 2), *Kcnq4* (potassium voltage-gated channel subfamily KQT member 4), *Ocm* (oncomodulin-1).



Supplementary Figure 8

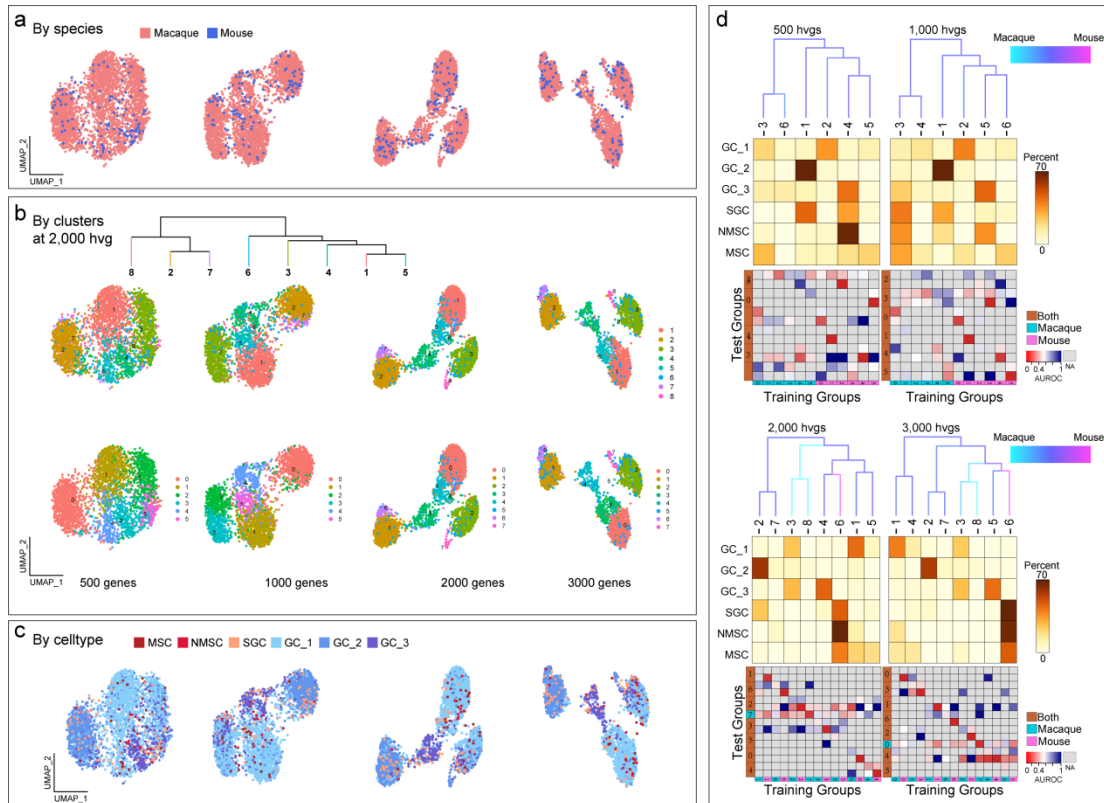
Sub-cluster analysis of mouse glial cells, related to Figure 3.

a UMAP (Uniform Manifold Approximation and Projection) representation of the re-clustering of mouse glial cells by canonical correlation analysis (CCA)-based integration, which reveals three subtypes.

b Violin plots illustrate the expression levels of pan-glial cell markers (*MBP*, *PLP1*, *SOX10*, *MOBP*) and subtype-specific markers in satellite glial cells (SGCs, *IGFBP6*, *GLUL*, *KCNJ10*, *SLC1A3*, *SLC6A15*, *GATA2*), myelinating Schwann cells (MSCs, *PMP22*, *PRX*, *NCMAP*), and non-myelinating Schwann cells (NMSCs, *SCN7A*). Full-name of these genes is: *MBP* (myelin basic protein), *PLP1* (myelin proteolipid protein), *SOX10* (transcription factor *SOX-10*), *MOBP* (myelin-associated oligodendrocyte basic protein), *IGFBP6* (insulin-like growth factor-binding protein 6), *GLUL* (glutamine synthetase), *KCNJ10* (ATP-sensitive inward rectifier potassium channel 10), *SLC1A3* (excitatory amino acid transporter 1), *SLC6A15* (sodium- and chloride-dependent neurotransmitter transporter *NTT73*), *GATA2* (endothelial transcription factor *GATA-2*), *PMP22* (peripheral myelin protein 22), *PRX* (periaxin), *NCMAP* (noncompact myelin-associated protein), *SCN7A* (sodium channel protein type 7 subunit alpha).

c UMAP representation of the re-clustering of macaque glial cells using Seurat-based joint canonical correlation analysis (CCA) embedding, revealing three subtypes.

d Violin plots showing the expression levels of pan-glial cell markers (*MBP*, *PLP1*, *SOX10*, *MOBP*, *GFAP*) and subtype-specific markers in SGCs (*IGFBP6*, *GLUL*, *KCNJ10*, *SLC1A3*, *SLC6A15*, *GATA2*), MSCs (*PMP22*, *PRX*, *NCMAP*), and NMSCs (*SCN7A*) that had been identified in mice and used for annotating macaque glial cell subtypes.



Supplementary Figure 9

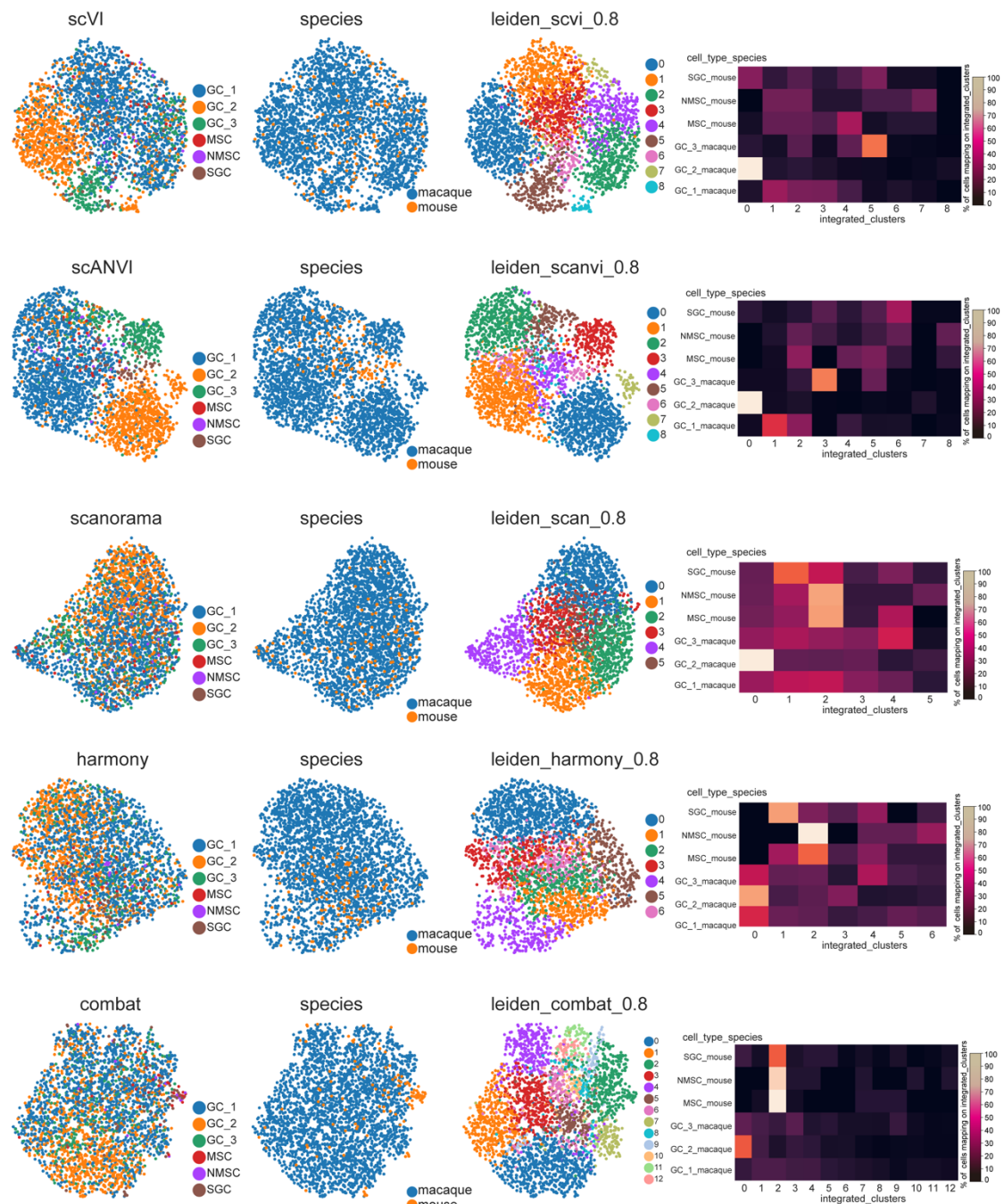
Integration of macaque and mouse glial cell data via the *Seurat* CCA-based pipeline by applying different numbers of highly variable genes (hvg), related to Figure 3.

a UMAP (Uniform Manifold Approximation and Projection) representations of single nuclei colored by species after canonical correlation analysis (CCA)-based integration, sequentially executed with differing numbers of hvgs, indicate unchanged species mixing.

b Top: hierarchical clustering of the average expression profiles was computed for the clusters identified in the integrated space (2000 hvgs), and colors in the terminal branches indicate the cluster number in the integrated data. Middle: UMAP representation of the integrations with colors computed with 2000 hvgs, which correspond to the color in the dendrogram. Bottom: UMAP plots of *Seurat* integrations with their original cluster number. Cells in the other UMAPs (500, 1000, and 3000 hvgs) are colored with the same colors used in the referenced integration (2000 hvgs), which indicates that the high-level structure of the data manifold is relatively stable when applying different numbers of hvgs.

c UMAP representation of single nuclei colored by glial cell subtypes after CCA-based integration running with 500, 1000, 2000, and 3000 hvgs.

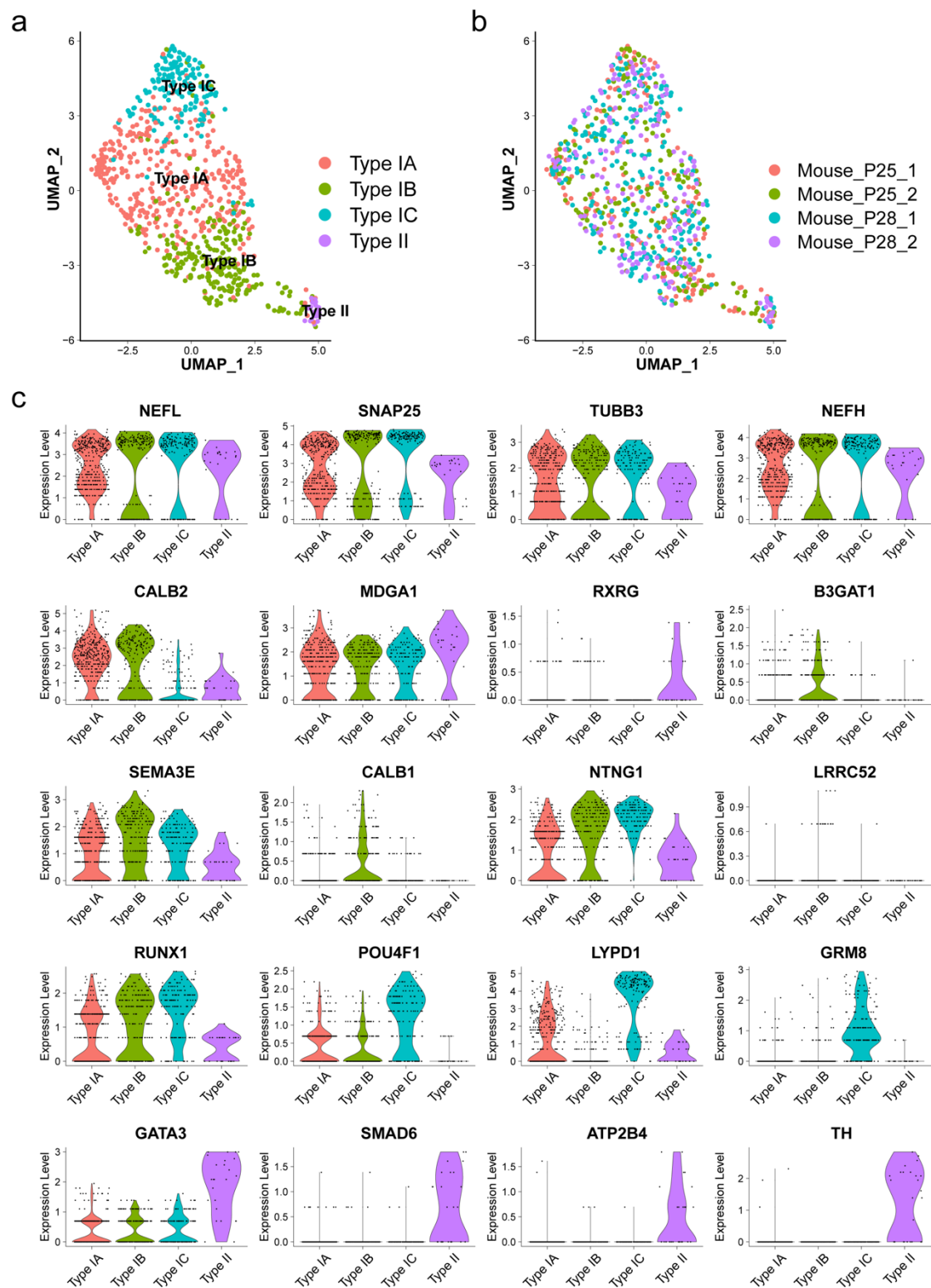
d Top panel: Hierarchical clustering of the average expression profiles of integrated glial cell clusters; branches are colored by species composition. Middle panel: Matrices displaying the percentage of cells in each integrated glial cluster (columns), with original subtypes grouped by species (rows). Bottom panel: Heatmaps illustrating the transcriptomic similarity of these integrated clusters within the original glial subtypes, determined by applying MetaNeighbor analysis.



Supplementary Figure 10
Integration of macaque and mouse glial cell data via multiple-integrative methods, related to Figure 3.

Left: UMAP (Uniform Manifold Approximation and Projection) representations of single nuclei colored by cell types, macaque and mouse species, and cluster finding resolution set of 0.8 computed by multiple integration algorithms used.

Right: heatmaps ordered in sequential columns displaying the cell mixings between Leiden algorithm-calculated clusters (x -axis) set by 0.8 resolution and original cell types (y -axis) within macaque and mouse species.



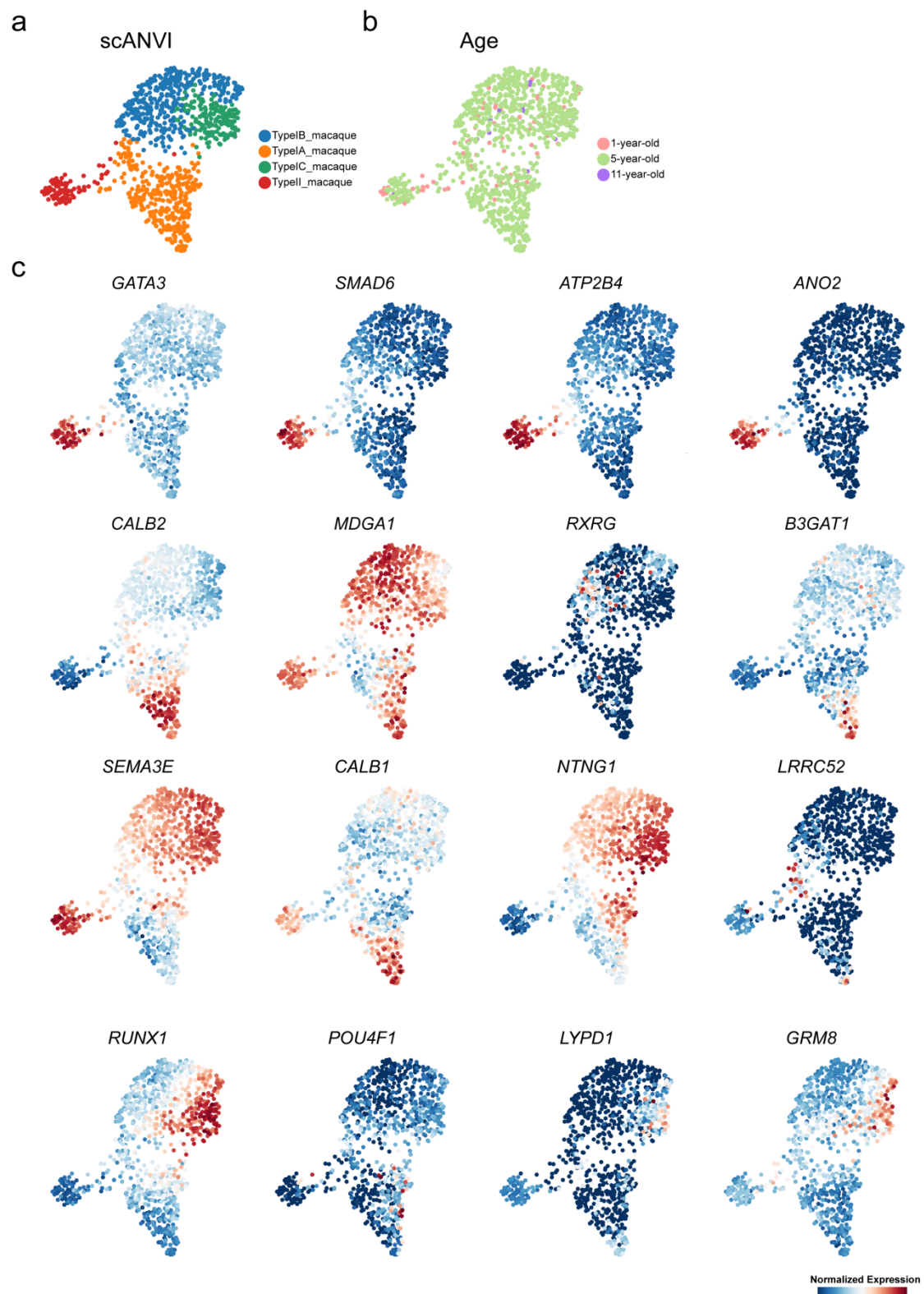
Supplementary Figure 11

Sub-clustering of mouse SGNs, related to Figure 4.

a, b UMAP (Uniform Manifold Approximation and Projection) representation shows the re-clustering of mouse SGNs from the postnatal days (P)25/28 cochlear snRNA datasets. 4 subtypes including three type I subtypes (IA, IB, IC) and one type II subtype are embedding and visualized. We used our own sequenced dataset of postnatal days(P) 25 mouse cochleae

and published P28 mouse cochleae (Zhenhang Xu et al. 2022, *Front Cell Neurosci.*) as a whole for SGN subtype analysis.

c Violin plots highlight the expression of pan-SGN markers (*NEFL*, *SNAP25*, *TUBB3*, *NEFH*) and subtype-specific markers in type IA (*CALB2*, *MDGA1*, *RXRG*, *B3GAT1*), type IB (*SEMA3E*, *CALB1*, *NTNG1*, *LRRC52*), type IC (*RUNX1*, *POU4F1*, *LYPD1*, *GRM8*), and type II (*GATA3*, *SMAD6*, *ATP2B4*, *TH*) SGNs. Full-name of these genes is: *NEFL* (neurofilament light polypeptide), *SNAP25* (synaptosomal-associated protein 25), *TUBB3* (tubulin beta-3 chain), *NEFH* (neurofilament heavy polypeptide), *CALB2* (calretinin), *MDGA1* (MAM domain-containing glycosylphosphatidylinositol anchor protein 1), *RXRG* (retinoic acid receptor RXR-gamma), *B3GAT1* (galactosylgalactosylxylosylprotein 3-beta-glucuronosyltransferase 1), *SEMA3E* (semaphorin-3E), *CALB1* (calbindin), *NTNG1* (netrin-G1), *LRRC52* (leucine-rich repeat-containing protein 52), *RUNX1* (Runt-related transcription factor 1), *POU4F1* (POU domain, class 4, transcription factor 1), *LYPD1* (Ly6/PLAUR domain-containing protein 1), *GRM8* (metabotropic glutamate receptor 8), *GATA3* (trans-acting T-cell-specific transcription factor GATA-3), *SMAD6* (mothers against decapentaplegic homolog 6), *ATP2B4* (plasma membrane calcium-transporting ATPase 4), *TH* (tyrosine 3-monooxygenase).



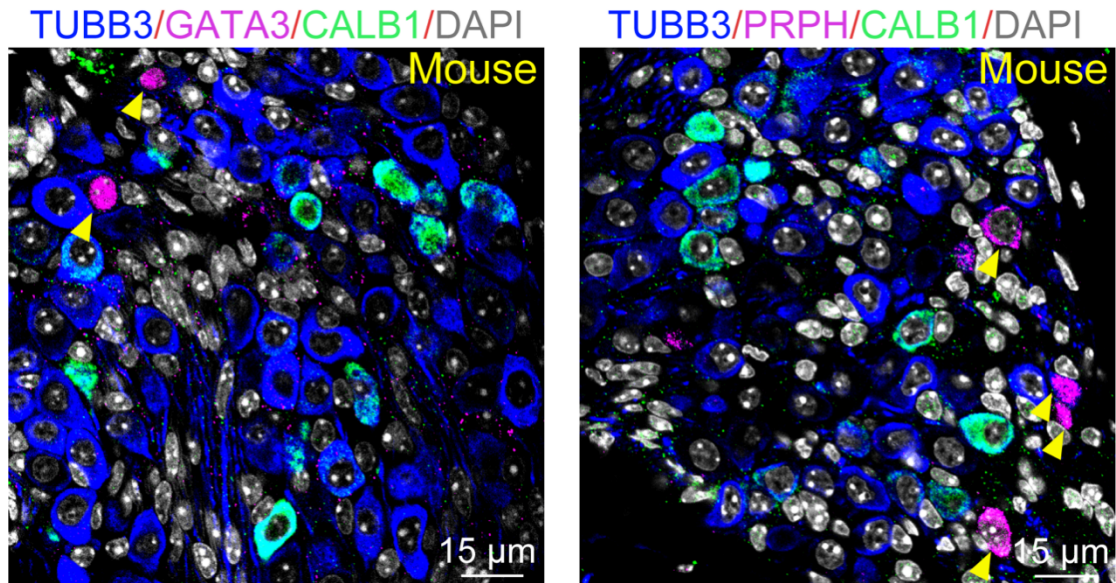
Supplementary Figure 12

SGNs in macaques are quite similar to mice, related to Figure 4.

a, b UMAP representation of the re-clustering of macaque SGNs into 4 subtypes.

c Feature plots illustrate the log-transformed normalized expression of subtype-specific markers in type II (*GATA3*, *SMAD6*, *ATP2B4*, *ANO2*), type IA (*CALB2*, *MDGA1*, *RXRG*,

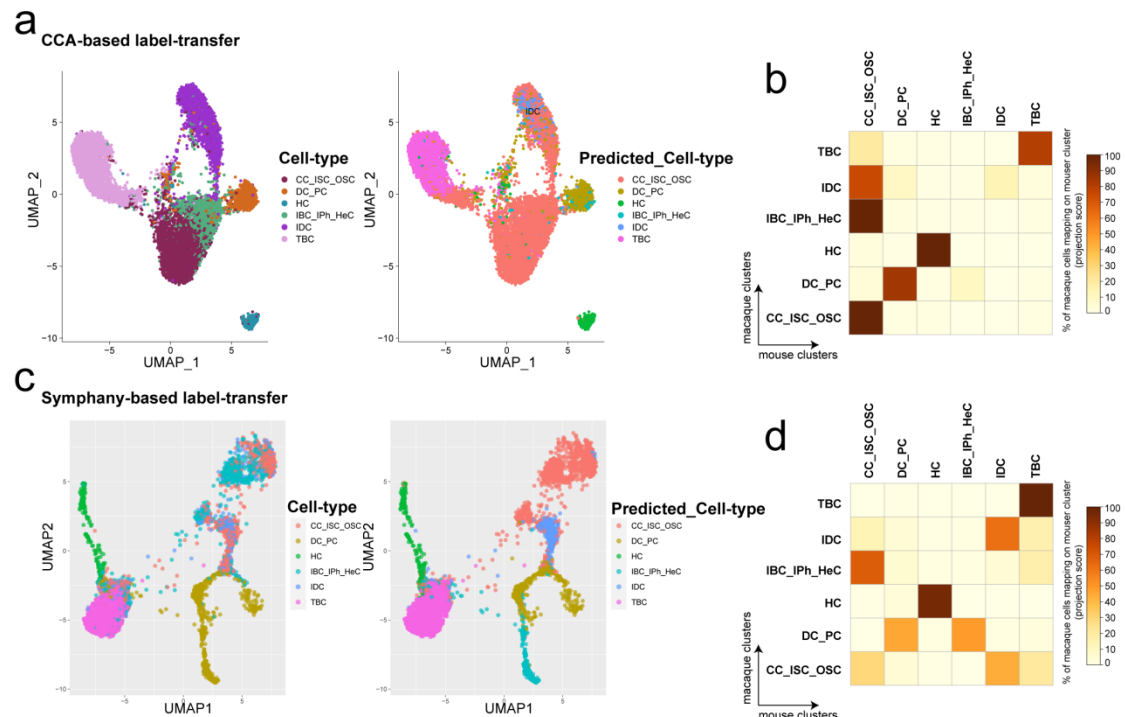
B3GAT1), type IB (*SEMA3E*, *CALB1*, *NTNG*, *LRRC52*), and type IC (*RUNX1*, *POU4F1*, *LYPD1*, *GRM8*) that were identified in mice and used for annotating macaque SGN subtypes.



Supplementary Figure 13

Immunofluorescence staining of CALB1 with GATA3 or with PRPH, related to Figure 4.

Representative confocal images of immunofluorescence cryosection staining demonstrating CALB1 associated with GATA3 or PRPH localization in postnatal days (P)21 mouse SGNs. Arrowheads highlighted positive signaling of GATA3 or PRPH, respectively.



Supplementary Figure 14

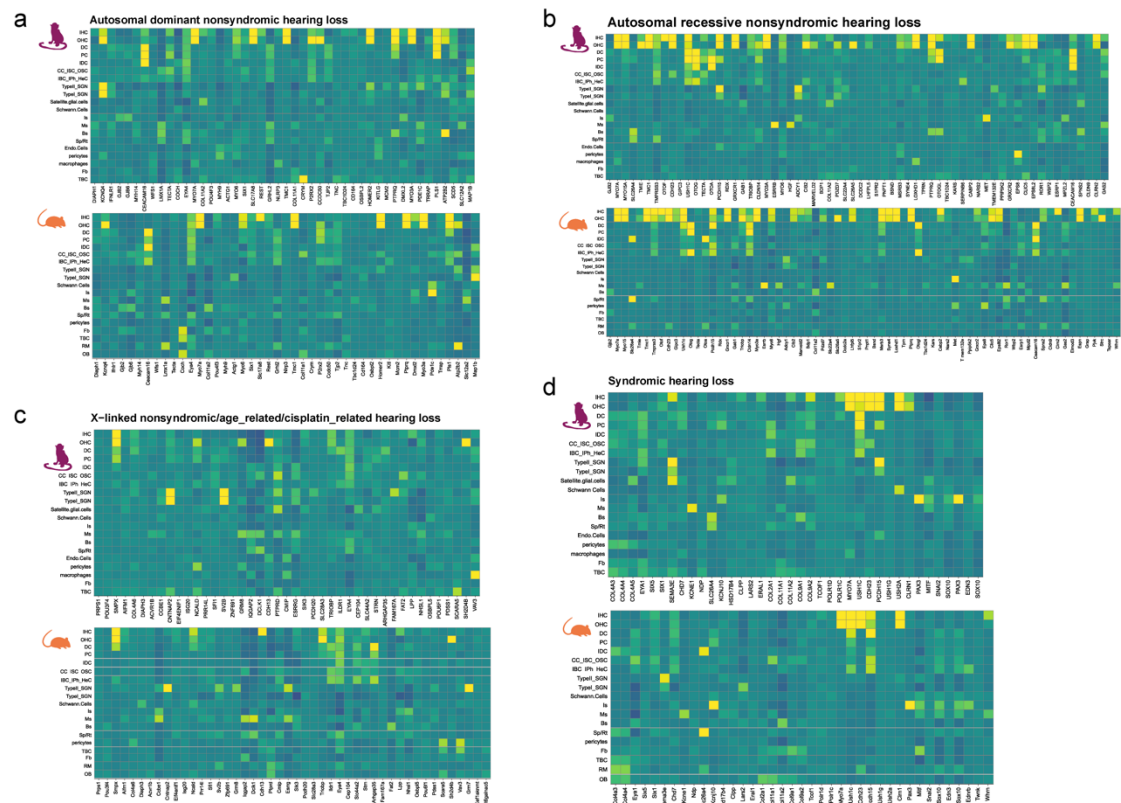
Macaque cochlear epithelium dataset label-transfers into mouse counterpart by CCA- and Symphany-based integration, related to Figure 5.

a UMAP representation of the predicted cell types calculated by canonical correlation analysis (CCA) and its original cell types from macaque cochlear epithelium.

b Heat map showing the percentage of cells from macaque mapping onto reference mouse cochlear epithelium cell types.

c same as in (a), but the analysis was done by Symphany-based label-transfer.

d same as in (b), but the analysis was done by Symphany-based label-transfer.



Supplementary Figure 15
Disease map of hearing-loss associated genes across the entire cochlear cell types, related to Figure 6.
The heatmap plots the expression of all known deafness-related genes, which are divided into four categories: **a** dominant nonsyndromic hearing loss (Dominant NSHL), **b** recessive nonsyndromic hearing loss (Recessive NSHL), **c** X-linked nonsyndromic / age-related / cisplatin-related hearing loss (X-linked / age-related / cisplatin-related NSHL), and **d** syndromic hearing loss (SHL), across the cochlear cellular map and divided by species. The gene display is derived from the Hereditary Hearing Loss Homepage.

Supplementary Data

Supplementary Data 1 GO enrichment analysis of cochlear cell-types in mouse and macaque.

Supplementary Data 2 RNAscope probe list