Cell Reports, Volume *23*

Supplemental Information

Transcriptional Dynamics of Hair-Bundle Morphogenesis Revealed with CellTrails

Daniel C. Ellwanger, Mirko Scheibinger, Rachel A. Dumont, Peter G. Barr-Gillespie, and Stefan Heller

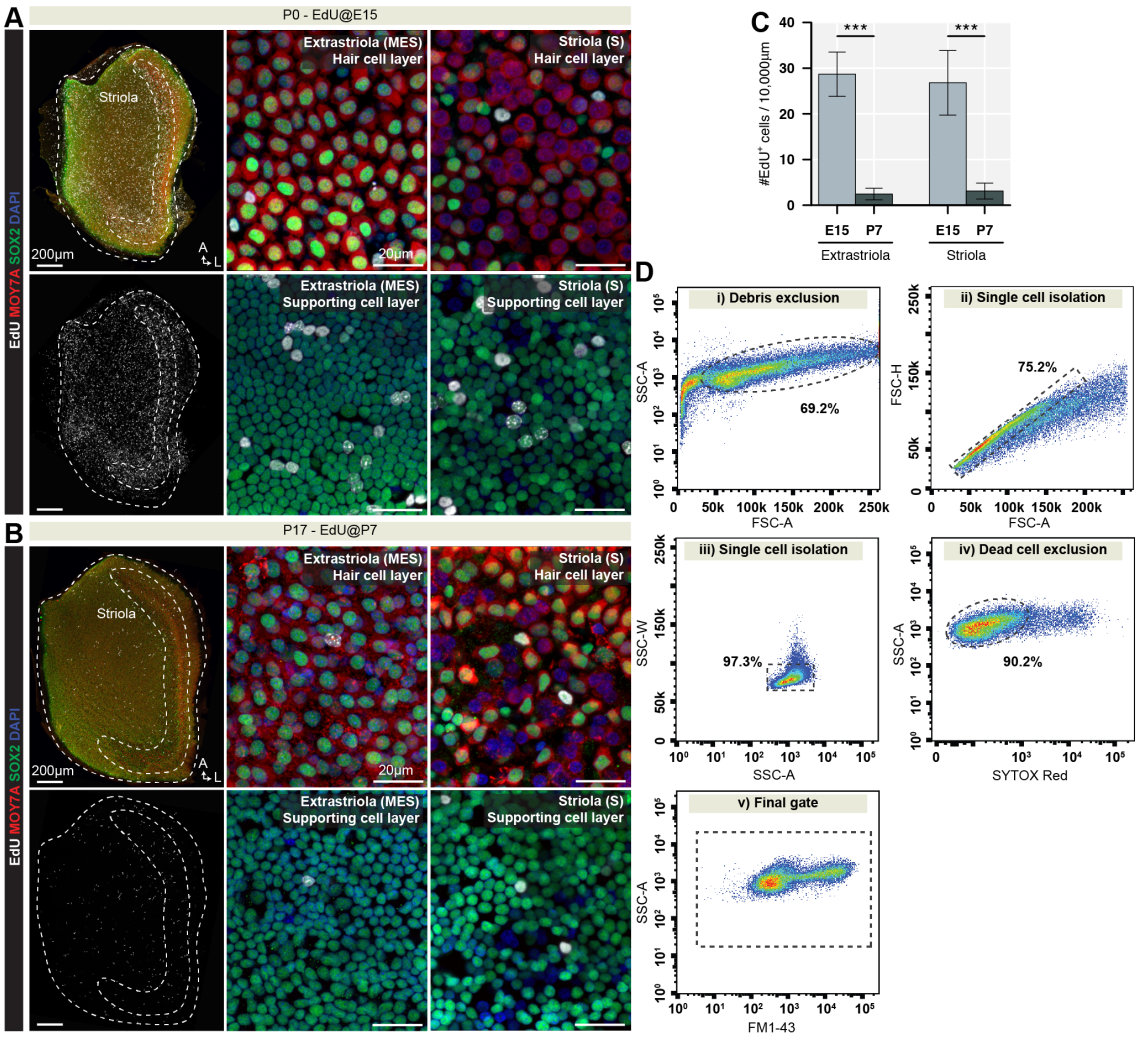
A
P0 - EdU@E15
Striola
Extrastriola (MES) Hair cell layer
Striola (S) Hair cell layer
Extrastriola (MES) Supporting cell layer
Striola (S) Supporting cell layer
200µm
20µm
EdU MOY7A SOX2 DAPI
B
P17 - EdU@P7
C
#EdU+ cells / 10,000µm
***
E15 P7
Extrastriola
Striola
D
i) Debris exclusion
69.2%
SSC-A
FSC-A
ii) Single cell isolation
75.2%
FSC-H
iii) Single cell isolation
97.3%
SSC-W
iv) Dead cell exclusion
90.2%
SYTOX Red
v) Final gate
FM1-43
E
n = 190 primer
NC
Serial dilution (log2)
Primer
CT
26.7
nd 4.2
F
Frequency
Coefficient of determination (R2)
Efficiency (%)
n = 185 primer (x > .90)
ATOH1
R² = 0.996
Efficiency = 104%
ATP2B2
R² = 0.999
Efficiency = 101%
CALB2
R² = 0.999
Efficiency = 107%
MYO7A
R² = 0.998
Efficiency = 107%
CT value
Input RNA (-log2 dilution)
G
Primer
Detection limit
LOD = 22.5
H
|M| = 2 × (11 × 96-well plates) = 11 × (185 gene × 96 cell panels)
Cells
Genes
I
CT value (GAPDH)
Cell
Panel
J
Frequency
m = 1,008 cells (95.5%)
Ct value (GAPDH)
K
Log2Ex
Panel
L
Frequency (thousands)
Log2Ex
M
CT ≤ LOD
Linear fit
Kernel density
High
Low
ΔΔCT
N
m = 1,008 cells, n = 183 genes
Panel

Figure S1. Specimen Analysis and Raw Data Processing. Related to Figure 1

(A-C) Utricle cell proliferation. 5-ethynyl-2'-deoxyuridine (EdU) was injected at E15 and utricles were examined at P7 (A), and EdU was injected at P7 and examined at P17 (B). (A, B) Immunolabeling with antibodies to hair cell marker MYO7A and supporting cell marker SOX2; nuclear DNA is labeled with 4,6-diamidino-2-phenylindole (DAPI). (C) Quantification of EdU^+ cells (per 100x100 μm^2 area) in extrastriolar and striolar regions of E15 and P7 chicken utricles. Error bars show the 95% confidence interval of the mean; *** Mann-Whitney test $P < 10^{-3}$.

(D) Illustration of the FACS gating strategy to filter viable single-cell events. Debris, dead cells, and cell aggregates were sequentially excluded (i-v). Indicated are gates and the percentages of filtered cells that propagated for each subsequent step from a representative experiment; SSC-A = side scatter area, SSC-W = side scatter width, FSC-A = forward scatter area, FSC-H = forward scatter height, and Sytox red, a dye excluded from living cells.

(E, F) Primer validation. C_T values for each primer pair were recorded for a serial two-fold 15-step dilution series of bulk RNA (NC = negative control). 185 suitable primer pairs had a linear relation between the generated C_T value and the log input RNA concentration ($R^2 > 0.9$ and efficiency > 0.9), such as illustrated with the examples of *ATOH1*, *CALB2*, *ATP2B2*, and *MYO7A*; error bars denote ± standard deviation of six replicates.

(G) Determination of the global limit of detection (LOD). The LOD was set to the 95% quantile of individual detection limits of the assayed primer pairs as identified by the dilution series.

(H) Missingness map. Shown is the pattern of non-detects (white points) of 195,360 measurements. The assay (185 primer pairs) was conducted on 11 panels, which are composed of two matching 96-well plates quantifying identical cells; black points indicate genes for which expression values were determined.

(I, J) Filtering of 1,008 cells based on the reference gene *GAPDH*. Red lines denote the inner and outer fence for 48 outliers (first and third quartile ± 1.5 × interquartile range).

(K, L) Distribution of Log2Ex transformed and normalized C_T values. Boxplots show similar Log2Ex value distributions (without non-detects) for each single panel. Histogram indicates a high frequency of non-detects in the data set.

(M) Relation between the probability of a primer producing non-detects and the mean normalized expression intensity (without non-detects) of the corresponding gene. Genes producing a higher number of transcripts have a lower probability for drop-outs.

(N) Hierarchical clustering of single cells colorized by individual panels suggests no significant batch effects.

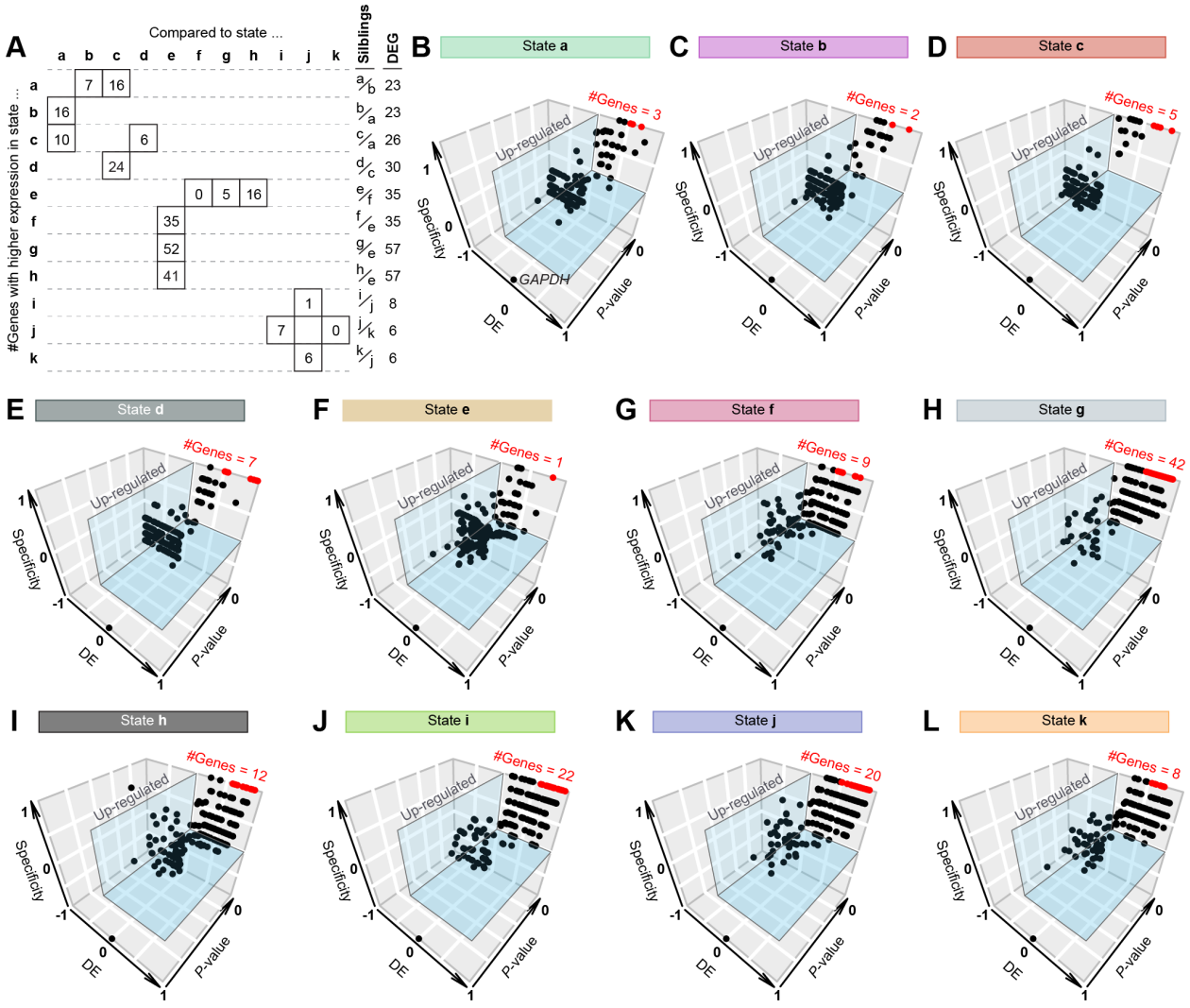

Figure S2. Gene Marker Detection. Related to Figure 1

(A) Post hoc analysis. The table represents the last iteration of the post hoc analysis generating the final set of states. It lists the number of genes significantly higher expressed in one state (row) compared to its neighbors (siblings) in the cluster dendrogram (column). For example, the closest neighbor to state *a* is state *b* in the cluster dendrogram. We found 23 differentially expressed genes (DEG) between both states, whereas state *a* has 7 genes significantly higher expressed compared to state *b*, and state *b* has 16 higher expressed genes than state *a*. As a consequence, states *a* and *b* were not joined and determined as distinct states.

(B-L) Multiple differential gene expression analysis. One state was compared to all other states simultaneously. Shown are the differential expression score (DE), specificity and Monte Carlo p of each gene for each state. Up-regulated genes are defined by DE > 0.5 and specificity > 0; specific state markers (red dots) have further specificity of 1 and $p < 10^{-3}$. As *GAPDH* was used for normalization, it has, as expected, DE = 0, specificity = -1 and *P* = 1 in all states.

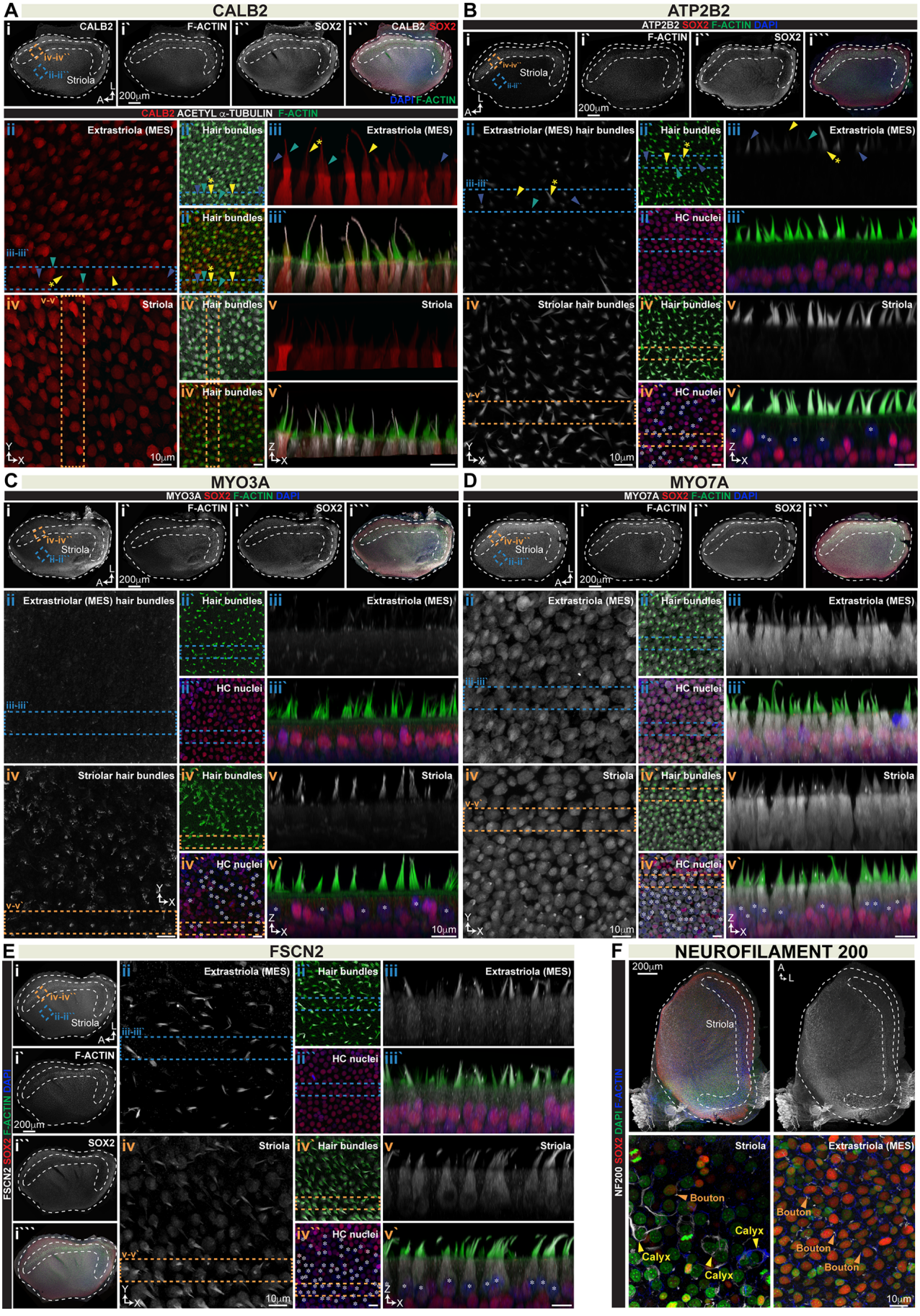
A
CALB2
CALB2 ACETYL α-TUBULIN F-ACTIN
Striola
Extrastriola (MES)
Hair bundles
HC nuclei
Striolar hair bundles
Extrastriolar (MES) hair bundles
B
ATP2B2
ATP2B2 SOX2 F-ACTIN DAPI
C
MYO3A
MYO3A SOX2 F-ACTIN DAPI
D
MYO7A
MYO7A SOX2 F-ACTIN DAPI
E
FSCN2
FSCN2 SOX2 F-ACTIN DAPI
F
NEUROFILAMENT 200
NF200 SOX2 DAPI F-ACTIN
Bouton
Calyx
200μm
10μm

Figure S3. Overview of Single-cell Protein Measurements. Related to Figures 4-6

(A-E) Immunolabeling of chicken E15 utricles. Nuclei were labeled with DAPI, stereocilia were visualized with phalloidin (F-ACTIN), kinocilia were labeled with antibodies to acetylated α-tubulin (ACETYL-α-TUBULIN), and supporting cells and type II hair cells were labeled with antibodies to SOX2. Images at different focal planes were recorded from the striola and medial extrastriolar regions as indicated (i-i'''). Entire volumes were rendered (z-stacks), which enabled the analysis of x-y projections (ii-ii'', iv-iv'') and x-z projections (iii-iii', v-v'); asterisks highlight nuclei of $SOX2^{-}$ maturing/mature type I hair cells. FSCN2 protein expression was not quantified. The arrowheads in A and B point to the same hair bundles that are labeled in Figure 4E and I.

(F) E15 utricles were immunolabeled with antibodies to the neural marker Neurofilament 200 kDa (NF200) and with antibodies to SOX2. DAPI was used to visualize nuclei. The E15 striolar epithelium harbors mature type I hair cells with calyx type terminals. Medial extrastriolar regions and the striola harbor mature type II hair cells, which are innervated by boutons.

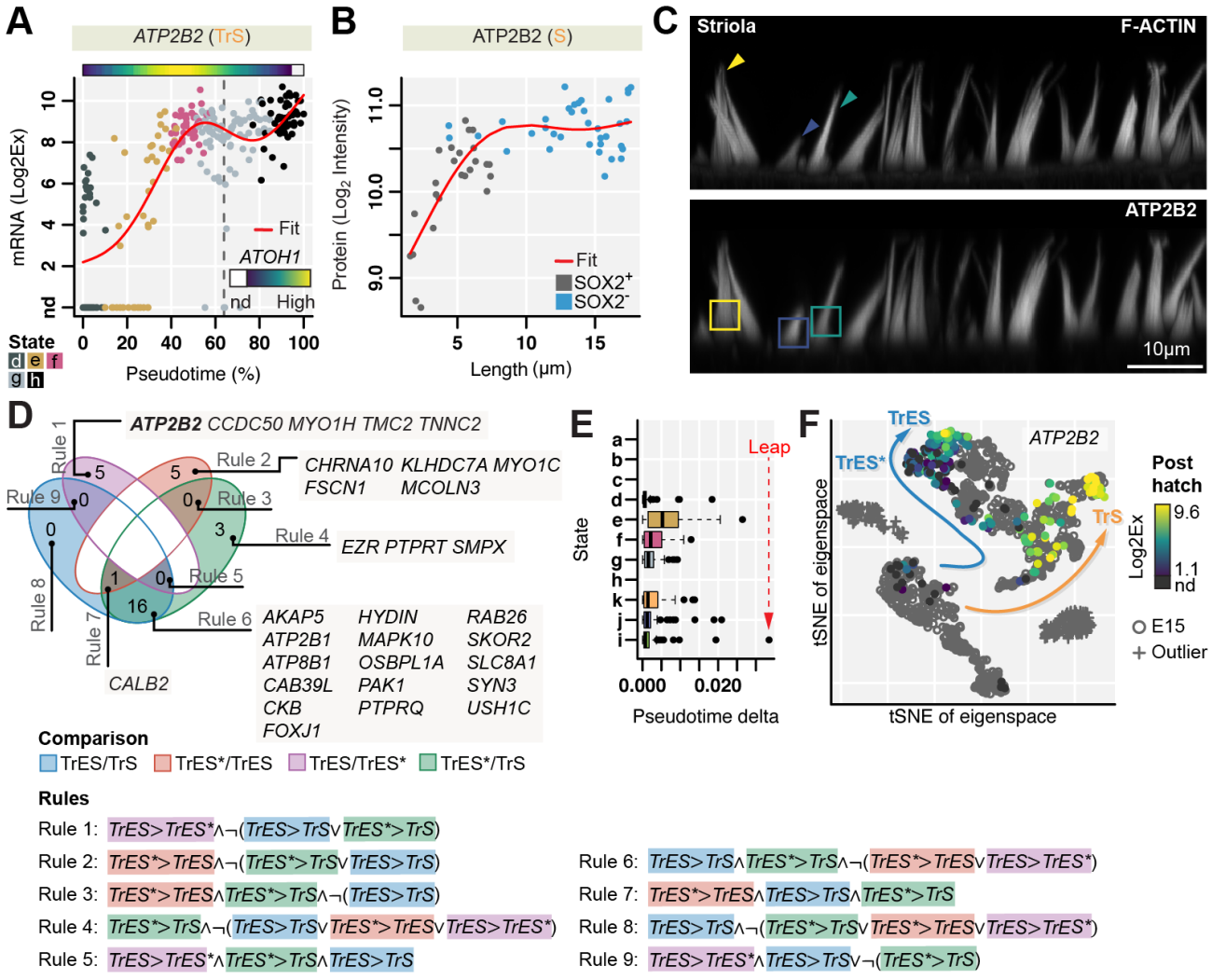

Figure S4. Striolar ATP2B2 Expression and Inter-Type II Hair Cell Comparison. Related to Figure 4

(A) *ATP2B2* expression of TrS. Cells are colorized by cluster (*d-h*), the fitted expression dynamic is shown in red; upper panel shows the corresponding *ATOH1* expression.

(B) ATP2B2 expression in the striola of E15 utricles. Quantification of ATP2B2 protein intensity along with bundle length measurements confirms up-regulation in developing type I hair cells (n = 61); lack of SOX2 expression confirms type I hair cell identity.

(C) ATP2B2 expression in short (blue), medium (green), and long (yellow) hair bundles.

(D) Venn diagram of gene sets with increased expression at the mature extrastriolar states. For example, *ATP2B2* is significantly higher expressed (Log2Ex fold-change > 2.5, Peto-Peto test $P < 10^{-3}$) at the terminal end state of TrES than TrES*, but not compared to TrS.

(E) Distributions of pseudotemporal cell distances in TrES. Shown are the boxplots of absolute differences between adjacent cells (pseudotime delta) within each state. The red arrow indicates a reasonable gap (strong outlier) along TrES in cluster *i*, indicating that this mature hair cell cluster may contain two sub-clusters. Whiskers extend to 1.5 times the interquartile range.

(F) Spectral embedding of embryonic and posthatch cells. The manifold was predicted to unfold in an 11-dimensional latent space (eigenspace). To visualize a two-dimensional approximation of the manifold, t-distributed stochastic

neighbor embedding (tSNE) was used. The branching trajectory structure is similar to E15 (TrS, TrES). Posthatch hair cells, shown as points colorized by *ATP2B2* expression intensity, aggregate at the terminal ends of the trajectory. Embryonic cells are shown as grey circles. Mature extrastriolar cells with higher and lower *ATP2B2* levels (TrES*) can be observed. Cells that were found unrelated to the trajectory (isolated clusters) are indicated by grey crosses; E15 outliers correspond to state *c*, posthatch outliers were found to be compromised hair cells based on distinct and significant upregulation (6.2-fold, Wilcoxon rank sum test $P < 5\times10^{-38}$) of *SH3GLB2* encoding endophilin 2, a protein that promotes mitophagy (Wang et al., 2016).

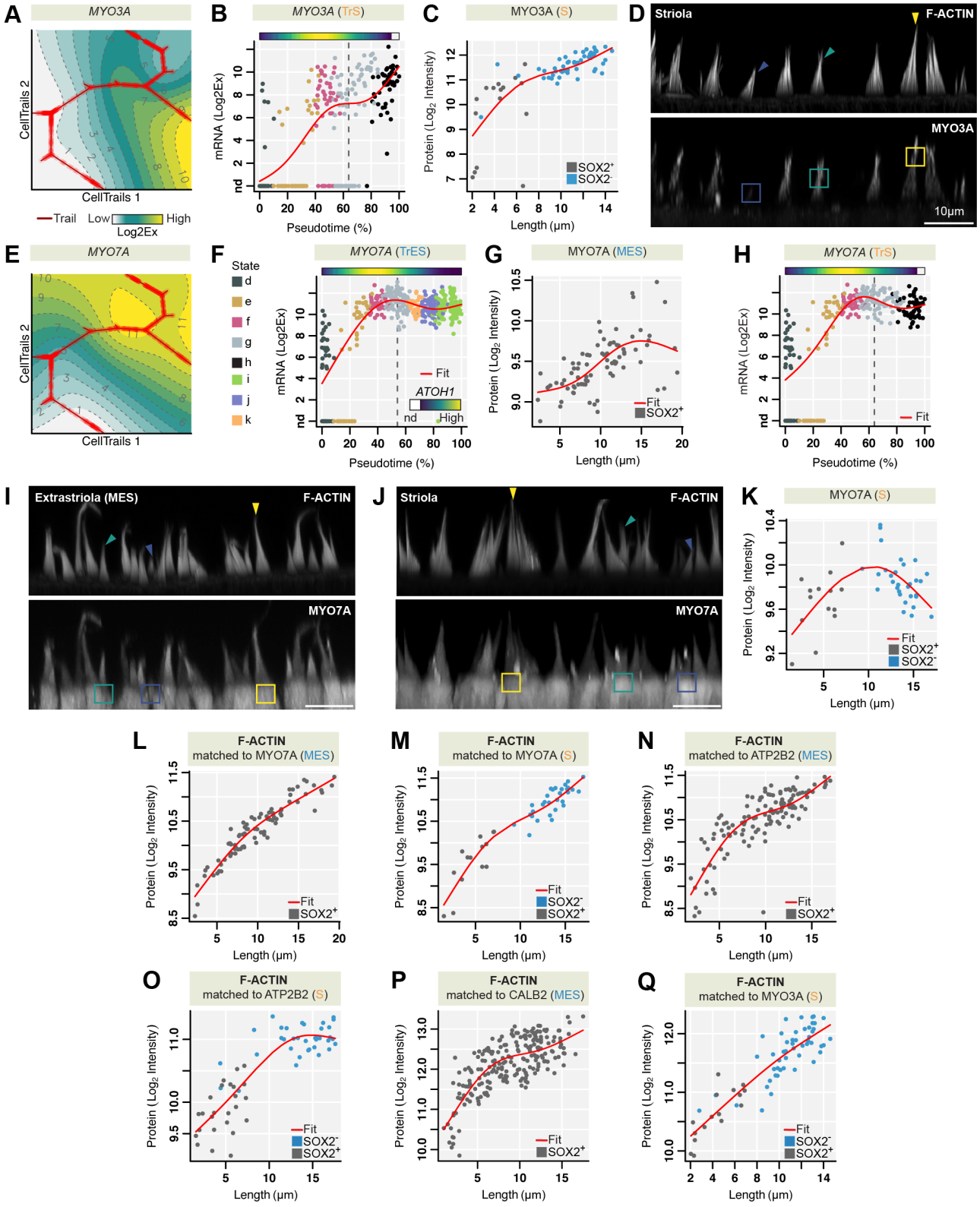

Figure S5. Transcript and Protein Expression of *MYO3A* and *MYO7A*. Related to Figure 6

(A-K) CellTrails maps and expression profiles of *MYO3A* and *MYO7A*. Immunolabeling qualitatively and quantitatively validated the inferred dynamics in hair cells from medial extrastriolar regions and hair cells from the striola (n = 66 MYO3A, n_{MES} = 83 MYO7A, $n_{striola}$ = 44 MYO7A measurements). Young hair cells with short bundles, irrespective of

future type, exhibit strong immunoreactivity for SOX2, which becomes downregulated in maturing type I hair cells (C, K). Arrows in immunolabeling images (D, I, J) indicate different bundle lengths (blue = short, green = medium, yellow = long).

(L-Q) For each single-cell target protein quantification (ATP2B2, CALB2, MYO7A, MYO3A), the corresponding F-ACTIN staining intensity was measured. The high positive correlation between bundle lengths and F-ACTIN levels indicates that unbiased and representative samples were obtained.

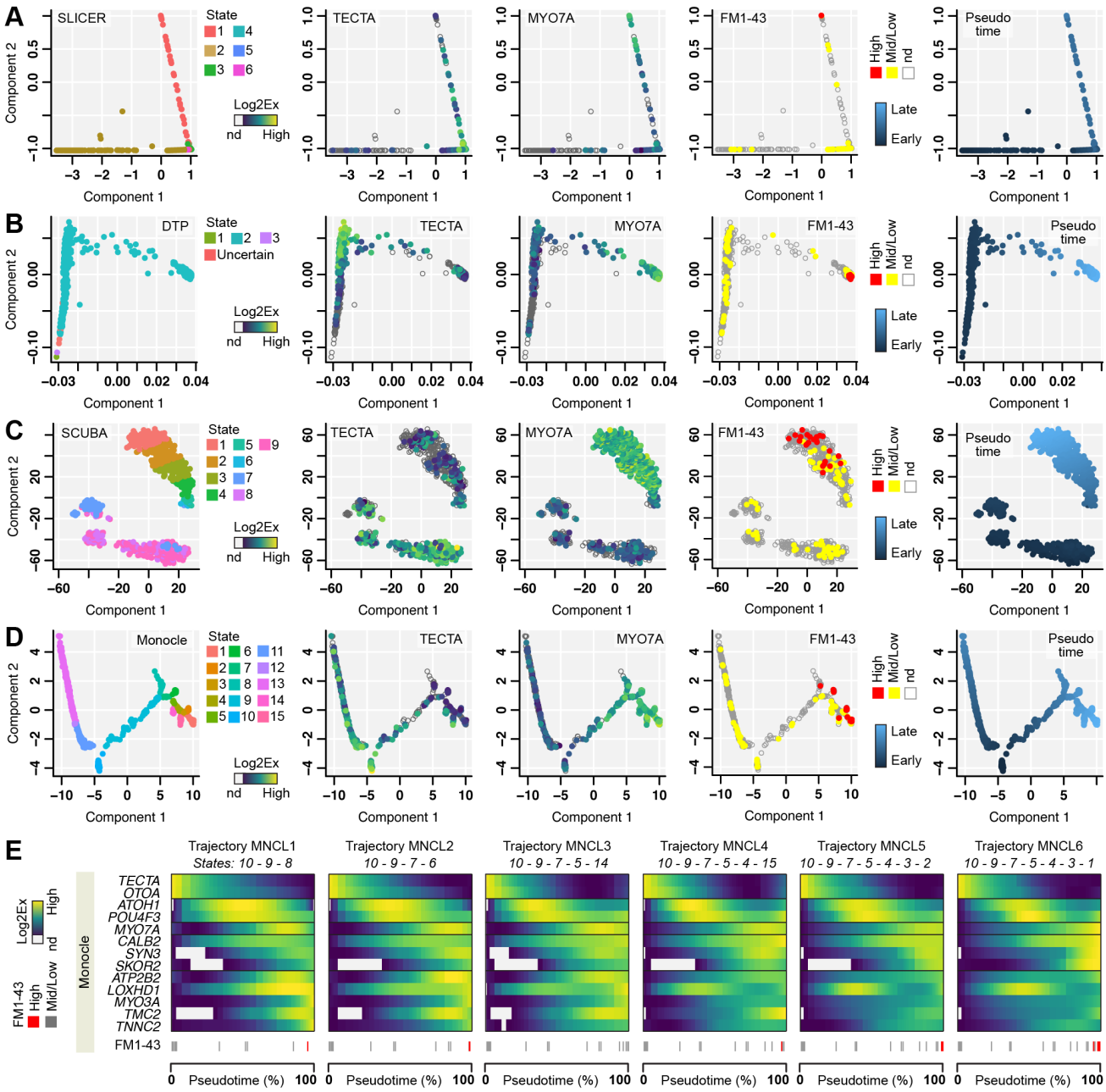

Figure S6. Use of Alternative Algorithms. Related to Figure 6

(A-D) Lower-dimensional representation of 1,008 cells computed by SLICER, DPT, SCUBA, and Monocle (version 2 with DDRTree). Shown are the states assigned by each algorithm, expression of the supporting cell marker *TECTA* and hair cell marker *MYO7A* (used to determine the directionality of the trajectory), FM1-43 dye labeling, and inferred pseudotime. All algorithms correctly order mature cells (FM1-43 high) at the terminal end or ends.

(E) Monocle predicts six differentiation endpoints, resulting in six different trails (sequence of states is concordant to (D)). The heatmaps show the expression dynamics of genes that are fully validated markers of striolar and extrastriolar hair cell development and were selected according to Figure 6A; heatmaps are scaled per gene and across trails.

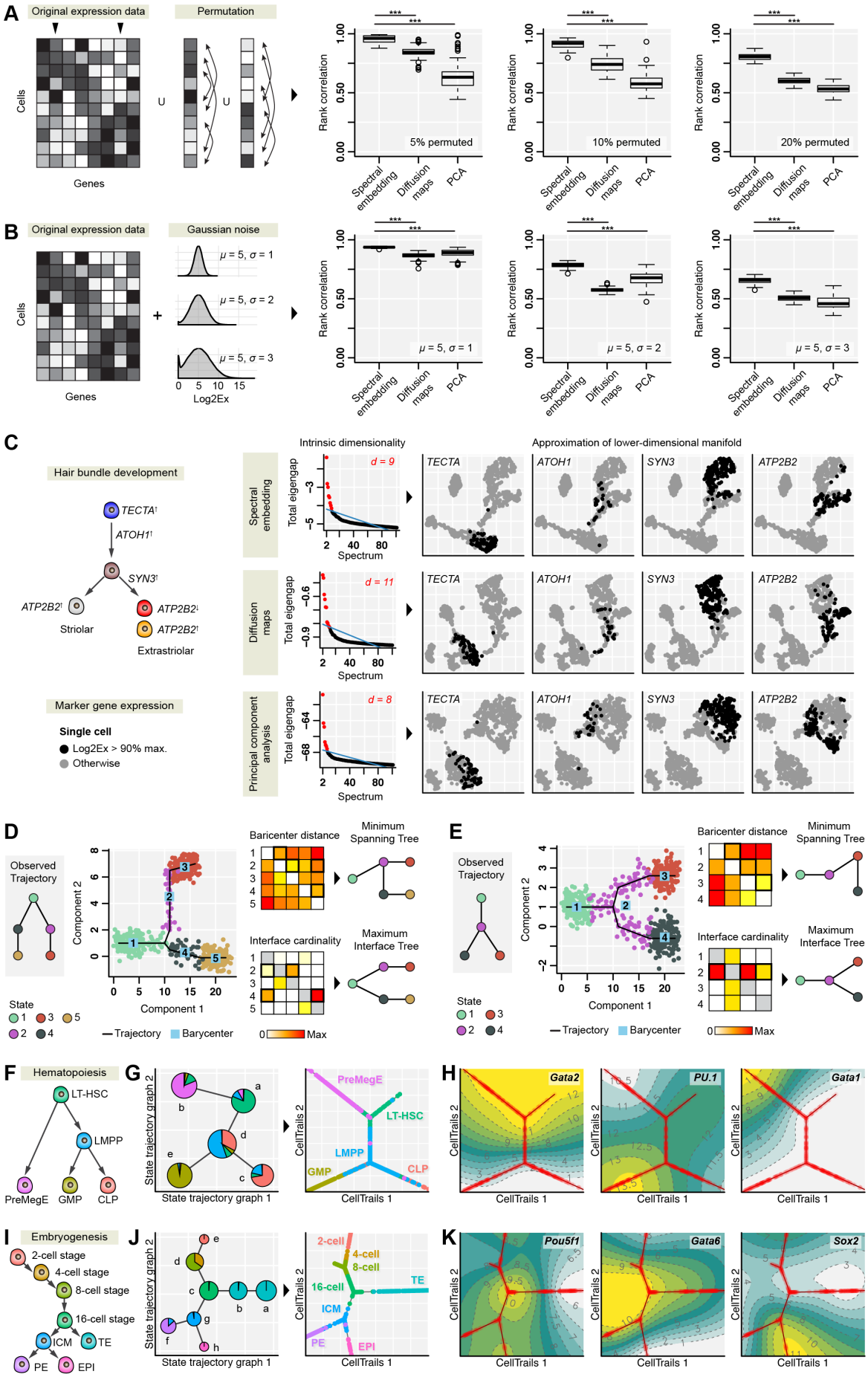
A
Original expression data
Permutation
Cells
Genes
U
Rank correlation
5% permuted
10% permuted
20% permuted
Spectral embedding
Diffusion maps
PCA
B
Original expression data
Gaussian noise
Cells
Genes
μ = 5, σ = 1
μ = 5, σ = 2
μ = 5, σ = 3
Log2Ex
C
Hair bundle development
TECTA
ATOH1
SYN3
ATP2B2
Striolar
Extrastriolar
Marker gene expression
Single cell
Log2Ex > 90% max.
Otherwise
Intrinsic dimensionality
Approximation of lower-dimensional manifold
Spectral embedding
Diffusion maps
Principal component analysis
Total eigengap
Spectrum
d = 9
d = 11
d = 8
D
Observed Trajectory
Component 1
Component 2
State
Trajectory
Barycenter
Baricenter distance
Minimum Spanning Tree
Interface cardinality
Maximum Interface Tree
Max
E
F
Hematopoiesis
LT-HSC
LMPP
PreMegE
GMP
CLP
G
State trajectory graph 1
State trajectory graph 2
H
Gata2
PU.1
Gata1
CellTrails 1
CellTrails 2
I
Embryogenesis
2-cell stage
4-cell stage
8-cell stage
16-cell stage
ICM
TE
PE
EPI
J
K
Pou5f1
Gata6
Sox2

Figure S7. Framework Assessment. Related to Figure 2

(A, B) Robustness analysis of spectral dimensionality reduction methods. Three noise levels were generated by sampling genes, shuffling the corresponding expression vectors, and by adding the permuted data to the original expression matrix, or by adding various degrees of Gaussian noise to the whole expression matrix (left panels). The absolute mean rank correlation between the top 10 reduced dimensions of the original data and the top 10 reduced dimensions of the data with a certain noise level was computed (right panels). The spectral embedding presented in this study shows significant higher robustness than diffusion maps, which were recently adapted for single-cell data analysis, and to the broadly used principal component analysis; n = 100 repeats per noise level; *** Wilcoxon rank sum test $P < 10^{-28}$.

(C) Visualization of the lower-dimensional manifolds derived by spectral methods. The intrinsic dimensionality was determined using the total eigengap (middle panels). The respective lower-dimensional manifolds were approximated and plotted in two-dimensions using tSNE (right panels); cells with high marker expression are indicated. Spectral embedding and diffusion maps show a branching trajectory towards striolar and extrastriolar hair cells; diffusion maps missed to identify the hair cell subtype (with higher levels of *ATP2B2*) at the terminal end in the extrastriolar branch. The left panel shows gene expression information for markers that were validated biologically in our study.

(D, E) Data was simulated for a trajectory with 5 states (D) as well as a trajectory with 4 states (E). In both examples, minimizing the total centroid distance failed to identify the correct trajectory structure, while maximizing the total interface cardinality (i.e., the number of neighboring cells between states) succeeded.

(F-K) CellTrails analysis of other datasets. (F, I) The left panels show the schematic of expected development hierarchies (Guo et al., 2010; Moignard et al., 2013). For each dataset the cell identity was annotated in the respective study (e.g., by FACS), but the age was latent. (G, J) CellTrails identified 5 and 8 trajectory states for hematopoiesis and embryogenesis, respectively. Pie charts represent the distribution of metadata per state; states were connected according to the predicted trajectory. CellTrails correctly reconstructs the branching linages and reveals lineage-specific expression dynamics of marker genes (H, K). For example, CellTrails maps delineate the mutual inhibitory expression pattern of the transcription factors *Gata1* and *PU.1* inducing hematopoietic progenitor linage commitment, as well as reveal continuous *Gata2* expression during PreMegE maturation and loss of *Gata2* expression during differentiation of LMPP-derived cell types; LT-HSC = long-term hematopoietic stem cell, PreMegE = pre-megakaryocyte erythroid, LMPP = lymphoid-primed multi-potent progenitor, GMP = granulocyte-monocyte progenitor, CLP = common lymphoid progenitor.