Cell Stem Cell
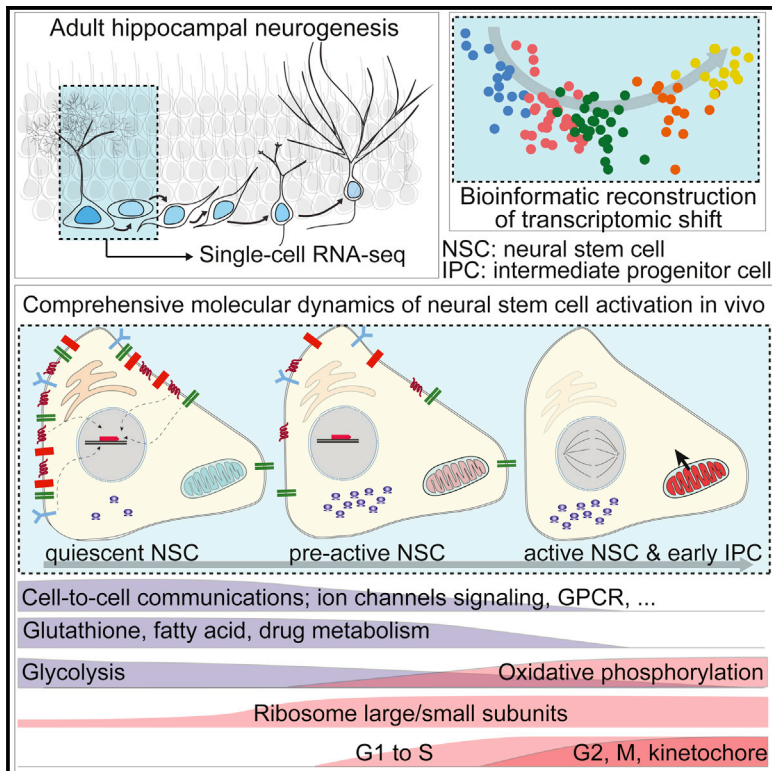
Single-Cell RNA-Seq with Waterfall Reveals Molecular Cascades underlying Adult Neurogenesis

Graphical Abstract

Highlights

- Single-cell transcriptomes from adult NSCs form a continuous linear trajectory

- Waterfall pipeline quantifies gene expression over continuous biological processes

- Adult NSC-specific metabolic pathways and active niche signal integration capacity

- Augmented translation capacity is the first hallmark of adult NSC activation

Authors

Jaehoon Shin, Daniel A. Berg, Yunhua Zhu, ..., Kimberly M. Christian, Guo-li Ming, Hongjun Song

Correspondence

shongju1@jhmi.edu

In Brief

In vivo molecular dynamics of adult stem cells have been elusive. Shin et al. used single-cell RNA-seq and a novel bioinformatic approach named Waterfall to reconstruct somatic stem cell dynamics with unprecedented temporal resolution. The genome-wide molecular transitions they identified suggest commonalities among different somatic stem cell systems.

Accession Numbers

GSE71485

CrossMark

CellPress

Single-Cell RNA-Seq with Waterfall Reveals Molecular Cascades underlying Adult Neurogenesis

Jaehoon Shin,[1,2] Daniel A. Berg,[2,3,7] Yunhua Zhu,[2,3] Joseph Y. Shin,[4] Juan Song,[2,3] Michael A. Bonaguidi,[2,3] Grigori Enikolopov,[8,9] David W. Nauen,[5] Kimberly M. Christian,[2,3] Guo-li Ming,[1,2,3,4,6] and Hongjun Song[1,2,3,4,*]

[1]Graduate Program in Cellular and Molecular Medicine
[2]Institute for Cell Engineering
[3]Department of Neurology
[4]The Solomon Snyder Department of Neuroscience
[5]Department of Pathology
[6]Department of Psychiatry and Behavioral Sciences
Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA
[7]Laboratory of Molecular Neurobiology, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, 17 117 Stockholm, Sweden
[8]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA
[9]Center for Developmental Genetics and Department of Anesthesiology, Stony Brook University, Stony Brook, NY 11794, USA
*Correspondence: shongju1@jhmi.edu
http://dx.doi.org/10.1016/j.stem.2015.07.013

SUMMARY

Somatic stem cells contribute to tissue ontogenesis, homeostasis, and regeneration through sequential processes. Systematic molecular analysis of stem cell behavior is challenging because classic approaches cannot resolve cellular heterogeneity or capture developmental dynamics. Here we provide a comprehensive resource of single-cell transcriptomes of adult hippocampal quiescent neural stem cells (qNSCs) and their immediate progeny. We further developed Waterfall, a bioinformatic pipeline, to statistically quantify singe-cell gene expression along a de novo reconstructed continuous developmental trajectory. Our study reveals molecular signatures of adult qNSCs, characterized by active niche signaling integration and low protein translation capacity. Our analyses further delineate molecular cascades underlying qNSC activation and neurogenesis initiation, exemplified by decreased extrinsic signaling capacity, primed translational machinery, and regulatory switches in transcription factors, metabolism, and energy sources. Our study reveals the molecular continuum underlying adult neurogenesis and illustrates how Waterfall can be used for single-cell omics analyses of various continuous biological processes.

INTRODUCTION

In discrete regions of the adult mammalian brain, quiescent neural stem cells (qNSCs) continuously generate new neurons through a recurrent process involving quiescent to active state transitions, cell-cycle entry, and neuronal fate specification (Ming and Song, 2011). Understanding molecular mechanisms underlying adult NSC regulation and neurogenesis will advance our knowledge of neural development and plasticity and enable new approaches for regenerative medicine and treatment of brain disorders. Mechanistic analysis of stem cell biology requires comprehensive quantification of molecular properties, such as gene expression. In contrast to traditional approaches targeting individual candidate genes, transcriptome profiling through RNA-sequencing (RNA-seq) provides an unbiased and quantitative proxy for molecular features of cellular states. Such a blueprint may reveal unexpected features of NSC biology, generate hypotheses for functional analysis, and lead to novel strategies to manipulate neurogenesis processes.

Classic approaches for molecular characterization of somatic stem cell behavior use population-based readouts at a few time points along development, which faces two major challenges: resolving cellular heterogeneity and capturing developmental dynamics. Adult stem cells constitute a minor population within complex tissues, intermingled with their progeny at different developmental stages and supporting cells. They switch among different states, such as quiescence and activation (Li and Clevers, 2010), and thus exhibit significant cellular and molecular differences even upon prospective isolation via fluorescence-activated cell sorting or genetic labeling (Codega et al., 2014; Lu et al., 2011). Furthermore, snapshots of molecular composition at selected time points are not sufficient to elucidate the dynamic nature of stem cell development.

Single-cell RNA-seq generates gene expression profiles at the resolution of an individual cell and has thus far revealed molecular profiles of cell types that were not previously recognized at the population level (Stegle et al., 2015). Single-cell RNA-seq has not yet been widely adopted for adult somatic stem cell studies due to technical difficulties in obtaining individual stem cells from complex tissues. Further, the stochastic nature of gene expression in individual cells (Muramoto et al., 2012; Novick and Weiner, 1957; Raj et al., 2006) may lead to overestimation of cellular heterogeneity and requires a new approach for statistical quantification. And for biological systems with only a few known markers, current approaches are not sufficient to map out a developmental trajectory at high resolution from single-cell datasets (Bendall et al., 2014; Trapnell et al., 2014).

Despite recent advances in acquiring snapshots of transcriptomes, epigenomes, and proteomes from individual cells, a remaining hurdle is the lack of methodology to identify
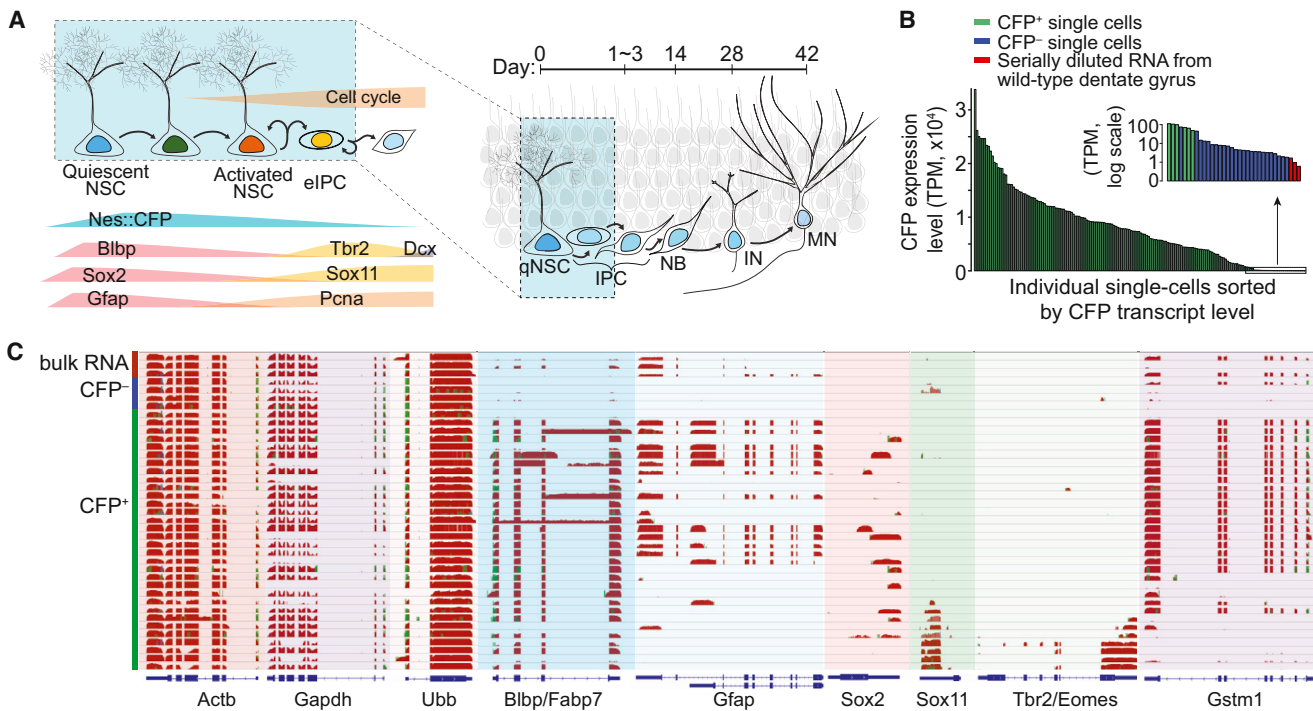
Figure 1. Single-Cell Transcriptomes of Adult Neural Stem Cells and Their Immediate Progeny

(A) A schematic diagram of the process of adult neurogenesis in the dentate gyrus of the mouse hippocampus. Once quiescent neural stem cells (qNSCs) become activated (aNSCs), they enter the cell cycle and generate early intermediate progenitor cells (eIPCs), which in turn give rise to neuroblasts (NB), immature neurons (IN), and then mature neurons (MN). Area highlighted with blue background indicates cell types fluorescently labeled in adult *Nestin-CFP^nuc* animals.

(B) Expression levels of transcript encoding CFP in each single-cell and diluted whole-dentate RNA samples (TPM, transcripts per million). Inset: enlarged view of CFP transcript levels in logarithmic scale of samples with low abundance of CFP transcript.

(C) Representative coverage profile of diluted total RNA from the whole dentate gyrus, CFP⁻ individual cells, and CFP⁺ individual cells at selected genomic loci, including house-keeping genes (β-Actin/Actb, Gapdh, and Ubiquitin B/Ubb), known NSC markers (Blbp/Fabp7, Gfap, and Sox2), known IPC markers (Sox11 and Tbr2/Eomes), and a potential new NPC marker (Gstm1).

See also Figure S1.

molecular state transitions over a developmental continuum. Cells are destroyed during acquisition of omic data, so the same cell can't be tracked over time. Here we developed a conceptually different approach, analogous to the "shot-gun" method used in the human genome project characterized by parallel sequencing and bioinformatic reconstruction. We focused on the narrow time window of adult qNSC activation and neurogenesis initiation. Using the *Nestin-CFP^nuc* transgenic genetic labeling system, we produced single-cell transcriptomes from a mixed population of precursor cells at different developmental stages. We then developed a bioinformatic pipeline named Waterfall to reconstruct continuous biological processes at single-cell resolution using adult neurogenesis as our model, and we applied this methodology to other stem cell datasets.

RESULTS

Single-Cell RNA-Seq of Neural Precursor Cells from the Adult Mouse Dentate Gyrus

In the adult dentate gyrus, radial-glia-like qNSCs give rise to new neurons via a sequential process of activation, proliferation, and generation of intermediate precursor cells (IPCs; Figure 1A) (Ming and Song, 2011). To elucidate detailed molecular dynamics during initial phases of adult neurogenesis in vivo, we

used a transgenic mouse line that expresses nucleus-localized CFP under the Nestin regulatory elements (*Nes-CFP^nuc*) (Encinas et al., 2006), in which CFP proteins carry over from adult qNSCs to their immediate progeny (collectively named NPCs; Figure S1A). The SMART-seq protocol (Ramsköld et al., 2012) was modified by the addition of a DNase I treatment step to remove genomic DNA for single-cell cDNA amplification (Figure S1A). In total, we performed single-cell RNA-seq for 142 CFP^nuc+ and 26 CFP^nuc− single cells (Table S1). Total RNA from wild-type adult mouse dentate gyri was serially diluted to 3 pg and processed in parallel for comparison.

We achieved, on average, 87% mapping onto annotated genes (Table S1). Sequencing reads were evenly distributed throughout the whole span of transcripts with 3′ bias comparable to recent studies (see Supplemental Experimental Procedures). Correlation analyses of RNA samples from different batches indicated minimal technical fluctuation during cDNA amplification or across batches compared to significant biological heterogeneity among single-cell transcriptomes (Figure S1B; Table S1). Universally expressed genes, such as β-Actin/Actb, Gapdh, or Ubiquitin B/Ubb, showed even expression patterns across all individual cells, whereas known NSC markers Gfap and Sox2, or early IPC (eIPC) markers Tbr2/Eomes and Sox11, were expressed in subsets of cells (Figure 1C; Table S2).
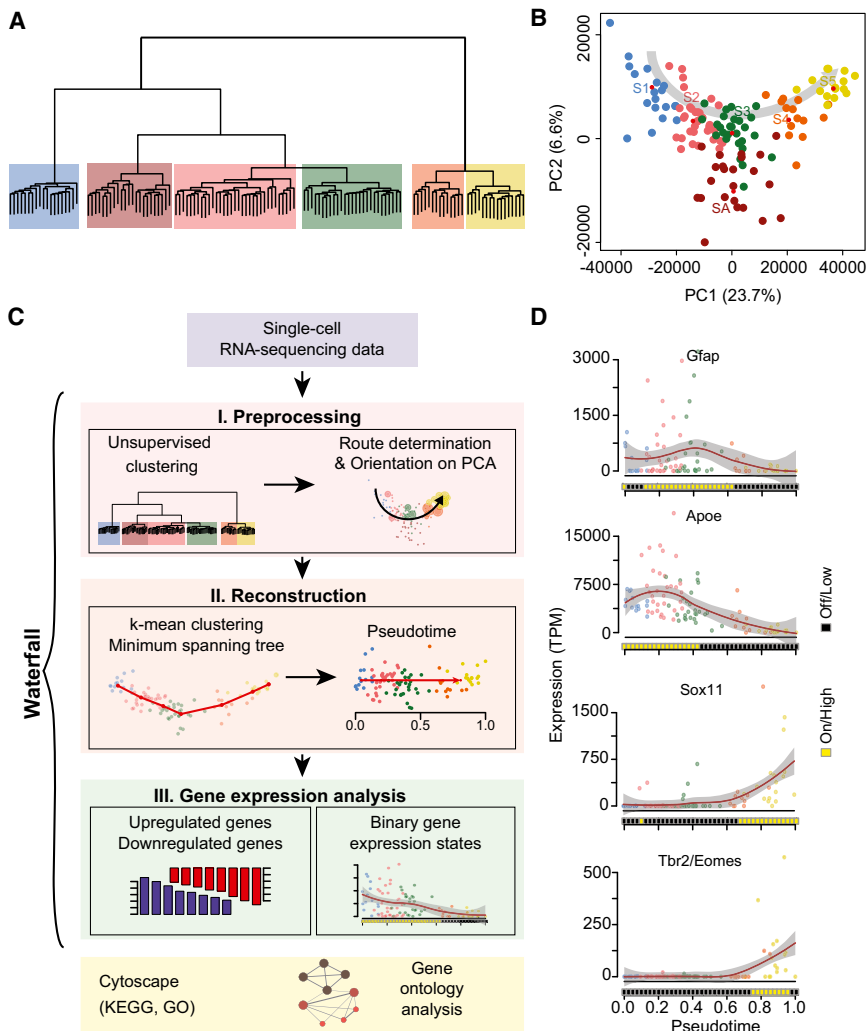
Figure 2. Waterfall for Analyzing Single-Cell Data from a Continuous In Vivo Process

(A) Unsupervised clustering analysis of CFP[nuc+] NPCs resulting in two super-groups with six sub-groups. Different groups are color-coded in the same fashion in this figure and across all other figures.

(B) Principal component analysis (PCA) plot shows one of the possible linear trajectories of different groups with the exception of SA.

(C) A schematic diagram of multiple components and workflow of Waterfall. Waterfall is a full range of algorithms for processing multi-dimensional single-cell datasets derived from continuous biological processes. Please see Supplemental Experimental Procedures for more information and Waterfall analyses of other biological systems.

(D) Representative expression profiles of marker genes of adult neurogenesis. Each data point represents the gene expression level of a single cell with a color scheme following that of Figure 2B. Data points are fitted with local polynomial regression fitting (red lines) with 95% confidence interval (gray area). HMM-predicted underlying states are represented as black and yellow squares on the bottom of the graphs.

See also Figures S2 and S3

The *Nes-CFP^nuc^* reporting system also labeled a small percentage of non-NPCs in the adult dentate gyrus (Figure S1) (Encinas et al., 2011). CFP transcript levels were multiple orders of magnitude higher in CFP[nuc+] cells compared to CFP[nuc−] cells or diluted dentate RNA (Figure 1B). We excluded cells that exhibited transcriptomic profiles markedly different from the majority of the CFP[nuc+] population or those that were were clearly identifiable as non-NPCs, such as oligodendrocyte progenitor cells or pericytes (Figure S1C). By differential expression analysis, we identified the top 35 genes enriched in CFP[nuc+] cells, which included known NSC markers Blbp, Spot14/Thrsp, Sox9, and GLAST/Slc1a3 (Table S2 and Table S3). In total, 31 out of 35 top genes exhibited subgranular zone (SGZ)-enriched expression patterns and/or were known NPC genes (p = 6.1 × 10^{-40}; hypergeometric test; Table S3). These results provided an initial validation of our approach.

Waterfall: Analyzing Single-Cell Datasets from Continuous In Vivo Processes

We next examined the whole-transcriptome dataset of individual CFP[nuc+] NPCs. Unsupervised hierarchical clustering analysis resulted in two super-groups with six subgroups (Figure 2A).

Notably, these six CFP[nuc+] groups were not clearly segregated on the plot of principal component analysis (PCA; Figure 2B), which was consistent over different batches of sequencing runs with multiple biological replicates (Figure S2A; Table S1). The continuous trajectory called for a new approach not relying on segmentation into a few groups of cell clusters. We could not use currently available single-cell analysis software, such as Monocle (Trapnell et al., 2010) or Wanderlust (Bendall et al., 2014), for our system due to the lack of sufficient prior information, such as temporal delineators or a robust set of specific markers (see Supplemental Experimental Procedures). We thus developed a more generally applicable pipeline of algorithms to perform unbiased statistical analyses of multi-dimensional single-cell datasets from continuous biological processes. We collectively named the suite of algorithms "Waterfall," which involves three steps: pre-processing, pseudotime reconstruction, and gene expression analysis (Figure 2C; see Experimental Procedures and Supplemental Experimental Procedures).

Pre-processing defined the trajectory of interest following dimensionality reduction of the data. Unsupervised clustering identified six clusters of cells (Figure 2A), which were then labeled S1–5 and SA based on their relative location in a PCA plot (Figure 2B). SA was recognized as a branch by the minimum spanning tree (MST) algorithm (see Supplemental Experimental Procedures). Although characterizing SA would be interesting (see Supplemental Experimental Procedures), we focused on the major neurogenic pathway in the current study. The expression profiles of a few known developmental genes were used

to orient the most probable trajectory of interest (Figures 2B and S2B).

To reconstruct the chronology, we first determined the most probable route of transcriptomic progression. We performed k-means clustering of single-cell transcriptomes on the PCA plot after excluding SA, then constructed an MST trajectory to connect cluster centers (Figure S3A). We then introduced "pseudotime" (Trapnell et al., 2014) to define the relative location of each cell on the MST trajectory (Figure S3A). Taking the Euclidian distance defined by the whole transcriptomic difference from each cell to the next in pseudotime, we found that the total path length reconstructed by Waterfall was significantly shorter than would result from random ordering of cells (see Supplemental Experimental Procedures). The pseudotime algorithm reconstructs molecular state transitions of a continuous process by quantifying the gradual divergence of single-cell transcriptomes individually, rather than as members of pre-classified groups.

For gene expression analysis, we developed an algorithm to determine the binary on/high or off/low expression state of each gene along pseudotime in an unbiased fashion using a hidden Markov model (HMM; Figure S3B). Gene expression at the single-cell level is highly stochastic and binary (Muramoto et al., 2012; Novick and Weiner, 1957; Raj et al., 2006). When analyzing continuous processes represented by single-cell transcriptomes in the absence of discrete groups, conventional statistical methods, such as arithmetic mean or t test, are not appropriate. We adopted HMM to statistically convert stochastic expression patterns of individual genes into binary on/high or off/low states (Figure S3B). The binary gene expression states were then shown as heatmaps, which quantified the molecular cascade over time (Figures 2C and 2D). To discover novel developmentally regulated genes, we correlated gene expression levels with pseudotime and subjected identified genes to gene ontology (GO) analyses (Figure 2C).

Validation for the Reconstructed Adult Neurogenesis Process

We validated molecular dynamics revealed by Waterfall at multiple levels. First, known NSC markers Gfap and Apoe, and eIPC markers Sox11 and Tbr2, showed non-overlapping expression states over developmental pseudotime (Figure 2D; Table S2). Second, we evaluated in vivo expression of Aldoc and Stmn1, which have not been previously studied in adult hippocampal neurogenesis. Over pseudotime, Aldoc was initially highly expressed (on/high state), then downregulated (off/low state), whereas Stmn1 was initially off, then upregulated (Figures 3A and 3C). In the hippocampal dentate SGZ of adult *Nes-GFP*[cyto] mice (Mignone et al., 2004), Aldoc[+]GFP[+] precursors were almost exclusively PCNA[−] qNSCs, with very few PCNA[+] active NSCs (aNSCs) or IPCs (Figure 3B). In contrast, Stmn1[+]GFP[+] precursors were mostly eIPCs and PCNA[+] aNSCs, but not PCNA[−] qNSCs (Figure 3D). Thus, Waterfall accurately predicted the in vivo expression dynamics of both known and unknown genes. Third, for functional validation, we explored the possibility of genetic labeling of a specific developmental stage based on Waterfall results. Hopx (Hop homeobox) was highly expressed in qNSCs but downregulated around the transition point from qNPC to aNSC (Figure 3E). Upon a single low-dose tamoxifen injection into a *Hopx-CreER*[T2] mouse line (Takeda et al.,

2011) crossed with an *mT/mG*[f/f] reporter line for clonal lineage-tracing (Bonaguidi et al., 2011), almost all labeled precursors at 3 days post injection were nestin[+]GFAP[+] qNSCs in the adult SGZ (Figure 3F). By 7 days, we were able to observe GFP-labeled clones that contained both NSCs and their progeny (Figure 3G), indicating self-renewal and differentiation, two hallmarks of NSCs.

Transcription Factor Expression during Adult NSC Activation and Neurogenesis

Waterfall allows an unbiased prediction of the relative chronological position of each individual cell and distribution of binary gene expression over the developmental trajectory. To delineate molecular cascades underlying adult qNSC activation and neurogenesis, we generated a list of the top 1,000 negatively correlated genes with pseudotime (DOWN[1000] genes; Spearman correlation coefficient < –0.13; Table S4), which represent qNSC-enriched genes downregulated during activation and neurogenesis, as well as the top 1,000 positively correlated genes with pseudotime (UP[1000] genes; Spearman correlation coefficient >0.20; Table S4), which represent newly activated genes during qNSC activation and early neurogenesis.

Out of these 2,000 genes, we initially focused on transcription factors (TFs). Most known TFs involved in adult neurogenesis were discovered by extrapolating findings from embryonic studies. In contrast, our database provides unbiased genome-wide profiles of TF expression. Systematic analyses of our dataset revealed a total of 41 downregulated TFs and 42 upregulated TFs during adult hippocampal neurogenesis (Figures 4A and S4; Table S5).

First, the set of dynamic TFs we identified included known regulators of adult NSCs and neurogenesis, which provided additional validation of our approach. Among DOWN TFs, Sox2, Sox9, Id3, nuclear receptor Nr2e1/Tlx, and Hes1 have been shown to regulate adult NSC maintenance and function. Among UP TFs, SoxC (Sox4 and Sox11), Foxg1, Tbr2, Insm1, Tcf12, and Nfib are critical in proliferative adult NPCs (Table S2).

Second, we identified multiple dynamic TFs that are regulators of embryonic neurogenesis but have not yet been studied in adult neurogenesis. DOWN TFs include homeobox protein Dbx2, nuclear glucocorticoid receptor Nr3c1, and Id4. UP TFs included chromatin protein Hmgb1 and proto-oncogene N-myc. During embryonic neurogenesis, these DOWN TFs inhibit the cell cycle (Nr3c1) or prevent premature differentiation (Id4), whereas UP TFs regulate progenitor proliferation (Hmgb1 and N-myc), suggesting conserved functions during embryonic and adult neurogenesis (Table S2).

Third, more than half of UP TFs and DOWN TFs are largely uncharacterized in the context of neurogenesis, but many of them are close paralogs or binding partners to other neurogenesis-related genes or have been implicated in other somatic stem cell systems. Examples include Hmgb1 paralogs (Hmgb2, Hmgb3, and Hmga1-rs1), SWI/SNF-related Brg1/Smarca4-associated factors (Smarcc1/Baf155 and Smarce1/Baf57), and Nfib paralogs (Nfia and Nfix). In addition, Mxd3, Zeb2, and ZT3/Zfp regulate adipocyte, melanocyte, and myogenic differentiation, respectively, whereas Tsc22d3/Gilz inhibits myogenic differentiation (Table S2). Hopx, which we found to mark qNSCs in the adult dentate gyrus (Figures 3E–3G), is expressed in
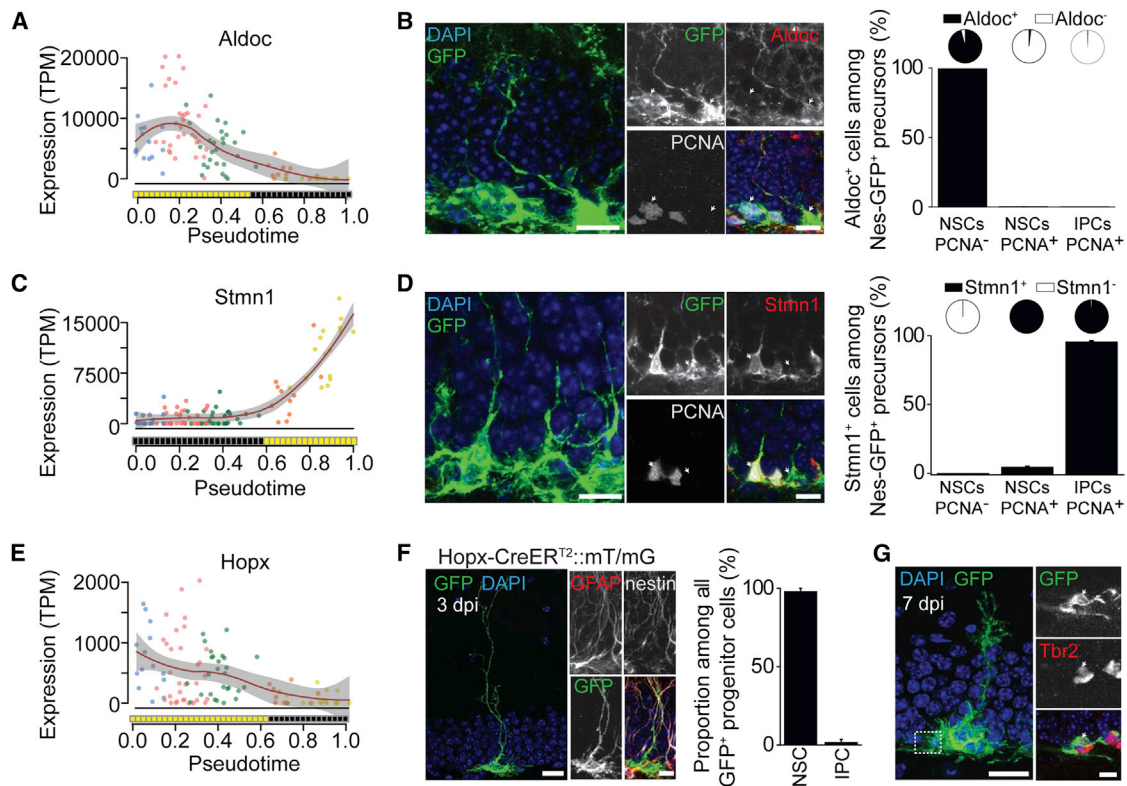
Figure 3. Validation for Waterfall Predictions for Early Adult Neurogenesis

(A–D) Validation of gene expression patterns and on/off binary states of Aldolase C (Aldoc, A) and Stmn1 (C) over pseudotime by immunohistology. Also shown are sample confocal images of GFP, cell proliferation marker PCNA, Aldoc, or Stmn1 in the dentate gyrus of adult *Nestin-GFP*cyto mice (left panels) and quantifications (right panels). Values represent mean ± SEM (n = 3 animals). The pie chart represents the proportion of Aldoc[+] or Stmn1[+] cells among each category of the GFP[+] progenitor population. Scale bars, 20 μm (left) and 10 μm (right).

(E–G) Validation of gene expression patterns and on/off binary states of Hopx over pseudotime (E) by genetic labeling and lineage tracing. Adult *Hopx-CreER*T2*::mT/mG* mice were injected with a single dose of tamoxifen and examined 3 days (F) or 7 days (G) later. Shown in (F) are sample confocal images of GFP, GFAP, Nestin, and DAPI. Also shown is the quantification of percentages of GFP[+] cells as NSCs or IPCs. Values represent mean ± SEM (n = 5 dentate gyri). Shown in (G) is an example of a labeled clone containing an NSC and multiple Tbr2[+] neuronal progeny. Scale bars, 20 μm (left) and 10 μm (right).

intestinal stem cells and a subset of multipotent hair follicle stem cells (Table S2).

Together, our single-cell transcriptome datasets are a rich resource of genome-wide dynamic expression profiles of TFs during adult neurogenesis. Many TFs that were previously uncharacterized in adult neurogenesis are known to be involved in embryonic neurogenesis or regulation of other somatic stem cells, suggesting shared biology among different stem cell systems and the potential utility of our resource for the general stem cell field.

Molecular Cascades underlying Adult qNSC Activation and Neurogenesis Initiation

The vast majority of UP[1000] and DOWN[1000] genes in our dataset were not TFs (Table S4). We investigated their characteristics from three perspectives: transition patterns along the developmental trajectory, cellular location of gene products, and biological function.

For transition patterns, plots of the top 150 genes each from UP[1000] and DOWN[1000] lists showed a wave of molecular activation or inactivation events over time, highlighting the sequential transition of gene expression during qNSC activation and neuro-

genesis (Figure 4B). To obtain biological insight into these transition patterns, we performed multiple gene ontology analyses. Strikingly, the predicted cellular localizations of protein products of UP[1000] genes and DOWN[1000] genes were drastically different. 51% of DOWN[1000] genes, as opposed to 20% of UP[1000] genes, encode proteins associated with the membrane (Figure 5A, $p = 2.6 \times 10^{-29}$). On the other hand, 58% of UP[1000] genes, as opposed to 20% of DOWN[1000] genes, encode proteins associated with the nucleus (Figure 5A, $p = 1.2 \times 10^{-36}$). Similar results were obtained using different thresholds for generating lists of UP and DOWN genes (Figure S5A).

Molecular Signatures of Adult qNSCs Revealed by DOWN[1000] Genes

Functional annotation of DOWN[1000] membrane genes revealed enrichment for ion or protein transport, cell communication, and cell adhesion (Figure 5B). Further classification identified proteins specific to the plasma membrane, ER, Golgi apparatus, and cytoplasmic vesicles (Figure S5B). KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway analysis revealed diverse functional entities involved in intra- and inter-cellular communication (Figure 5B). Specifically, genes related to Notch

signaling, GABAergic synapses, glutamatergic synapses, BMP pathways, the MAPK pathway, calcium, and cell adhesion were downregulated upon qNSC exit from quiescence (Figures 5B, 6A, 6B, and S6A). Electrophysiological recordings of *Nestin-GFP*cyto+ NSCs in acute slices from adult animals showed responses to both AMPA and NMDA, suggesting expression of functional receptors (Figure 6B). Each functional signaling pathway entity contained key genes that encode receptors, subunits, or downstream mediators (Figures 6A and 6B). Notably, many ligands for these receptors, including glutamate, GABA, Wnts, BDNF/neurotrophin, Jagged1, BMPs, FGFs, and Insulin/IGF2, are known to be present in the adult SGZ niche (Table S2), suggesting active signaling in qNSCs. While previous studies have examined each of these ligands and receptors in regulation of adult neurogenesis in isolation, our systematic genome-wide analyses unified disparate information and suggested a novel model wherein quiescent adult NSCs are not passive or dormant, but instead are actively integrating various niche signals. More surprisingly, qNSC activation was associated with decreased expression of genes involved in transducing local environment cues and pervasive downregulation of various signaling-pathway-related genes (Figure 6A). These results suggest that, once activated, adult NSCs shunt their capacity to respond to external regulation.

KEGG analysis of DOWN[1000] genes also revealed a shift in energy source and metabolism. First, multiple lipid-metabolism-related functional entities, including fatty acid degradation and sphingolipid metabolism, were enriched in qNSCs, but downregulated upon activation (Figure 5B). As previously reported (Knobloch et al., 2013), qNSCs exhibited the highest level of Spot14 (Figure S6B), which regulates lipid metabolism. qNSCs also maintained an active fatty acid degradation pathway (Acsl3, Acsl6, and Acsbg1; Figure S6B; Table S2). Second, pathway analysis consistently indicated glutathione metabolism and glycolysis as an adult qNSC characteristic, which was lost upon activation. Among glycolysis genes, aldolase A, aldolase C, and Ldhb decreased significantly, whereas most other glycolysis genes, including Gapdh, did not change during initiation of neurogenesis (Figure S6C; Table S2).

To validate results from analyses of the top 1,000 significantly downregulated genes, which contained a limited number of genes in each particular pathway, we performed analysis using all expressed genes and an independent functional annotation database wikipathway (Pico et al., 2008). Virtually identical results were obtained (Figure S7; Table S7). Together, analyses of downregulated genes provided novel insight into molecular signatures of adult qNSCs, including both intrinsic properties and regulation of intra- or inter-cellular signaling pathways.

Sequential Molecular Dynamics during Adult Neurogenesis Revealed by UP[1000] Genes

We next analyzed UP[1000] genes, which were nucleus-associated and/or related to the cell cycle, DNA/RNA metabolism, and chromosome organization (Figure 5A). Detailed analysis revealed pervasive activation of cell-cycle-related genes, ranging from cell-cycle-supporting genes, such as those involved in nucleotide synthesis, protein/RNA synthesis, and DNA fidelity controls (DNA repair and p53 signaling pathways), to genes directly involved in the cell cycle, such as those involved in DNA replica-

tion, the kinetochore complex, cyclin/cyclin-dependent kinases, or the cytosolic mitotic spindle (Figure 5B).

As opposed to downregulation of glycolysis-related genes, oxidative-phosphorylation-related genes were upregulated (Figure S6C). Specifically, in contrast to stable expression of earlier mitochondrial respiratory chain complexes (complex I, II, III, and IV), expression of subsequent complexes (complex V) increased over pseudotime, implying a gradual completion of the full electron transport chain during neurogenesis (Figure S6C).

The high resolution of Waterfall analyses revealed temporal relationships among genes in different functional groups. Cell-cycle checkpoint genes were sequentially activated following the known biological sequence of cell cycles: G1 to S transition, followed by G2 to M transition, and then chromosomal segregation, indicating that our pseudotime accurately reconstructs sequential biological events (Figure S6D). Initiation of the cell cycle preceded the major transcriptomic shift (Figure 6C). Notably, upregulation of genes encoding ribosomal subunits preceded the appearance of any cell-cycle checkpoint genes (Figures 6C and S6D), suggesting that priming of protein synthesis machinery may mark the G0 to G1 transition ahead of cell-cycle entry during adult qNSC activation.

Together, analyses of UP[1000] genes suggested that molecular dynamics of qNSC activation and initiation of neurogenesis are largely defined by priming of protein synthesis machinery, cell-cycle entry, activation of RNA and protein biogenesis, and a shift in energy metabolism from glycolysis to oxidative phosphorylation. An independent approach using a different functional annotation database showed similar results (Figure S7; Table S7).

Holistic Picture of Molecular Cascades underlying Adult Neurogenesis Initiation

Based on the molecular dynamics from qNSCs to aNSCs and then eIPCs, we have reconstructed sequential waves of biological events from single-cell RNA-seq data and Waterfall (Figure 7). The process begins with adult qNSCs downregulating TFs defining quiescence and decreasing competence for cell signaling (RTKs, GPCRs, neurotransmitter receptors, cytokines, and calcium). Concurrently, glycolysis, glutathione, and fatty acid metabolism begins to wane, while upregulation of protein translation capacity is the first marker of a pre-activation stage. As NSCs enter the cell cycle, oxidative phosphorylation becomes the primary energy source. Progression through the cell cycle accompanies a major decline in NSC metabolism (glutathione, fatty acid, and drug metabolism) and an increase in eIPC TFs. Finally, kinetochore and chromosomal segregation occurs in the first neurogenic progeny. Overall, the developmental trajectory is defined by a coordinated switch from a membrane-targeted to a nuclear-targeted transcriptome, suggesting a transition from qNSCs dominated by extrinsic signaling to eIPCs dominated by a pre-programmed intrinsic molecular cascade.

DISCUSSION

Understanding adult NSC behavior and neurogenesis requires quantification of molecular states along a continuous developmental process. In the current study we generated three major
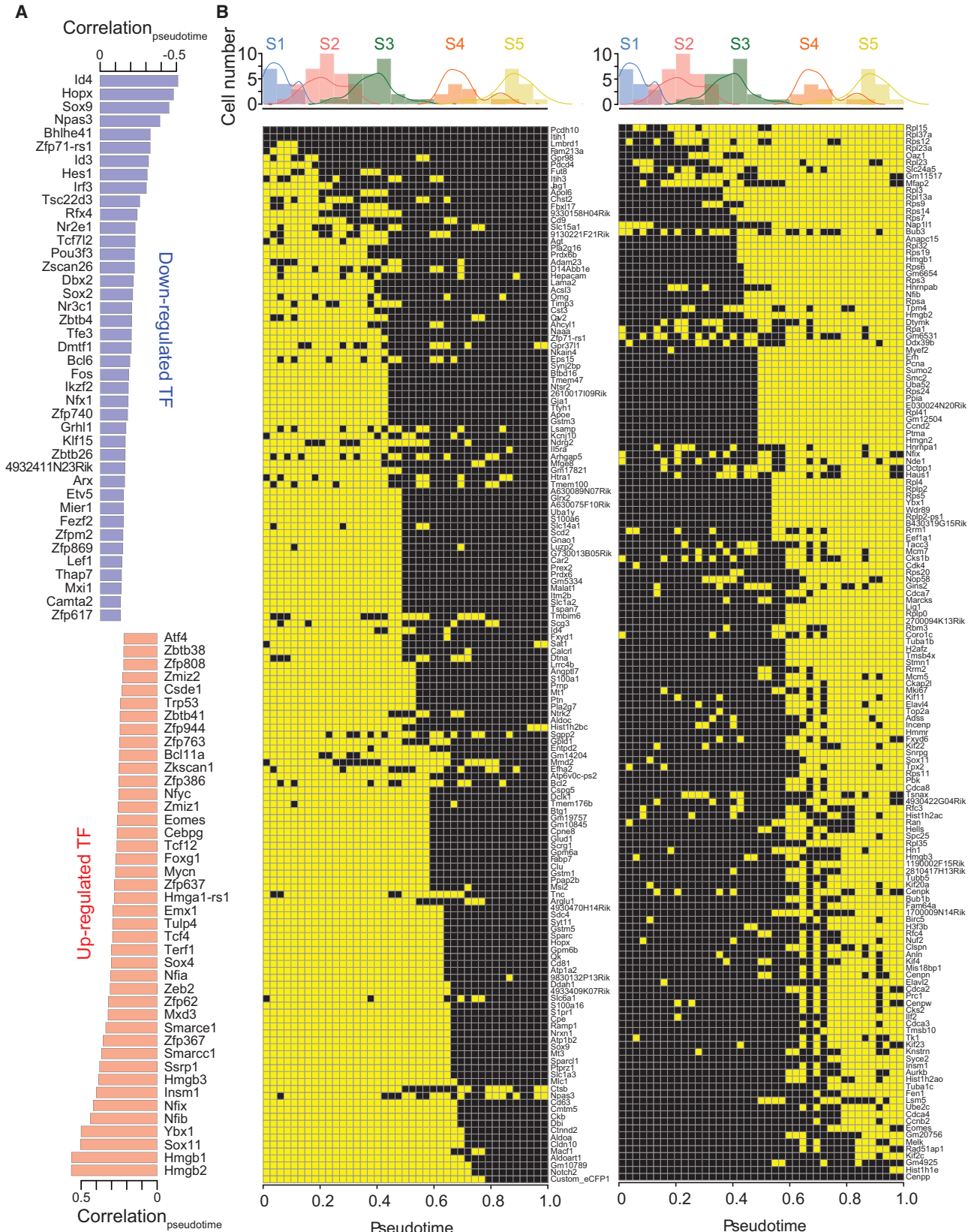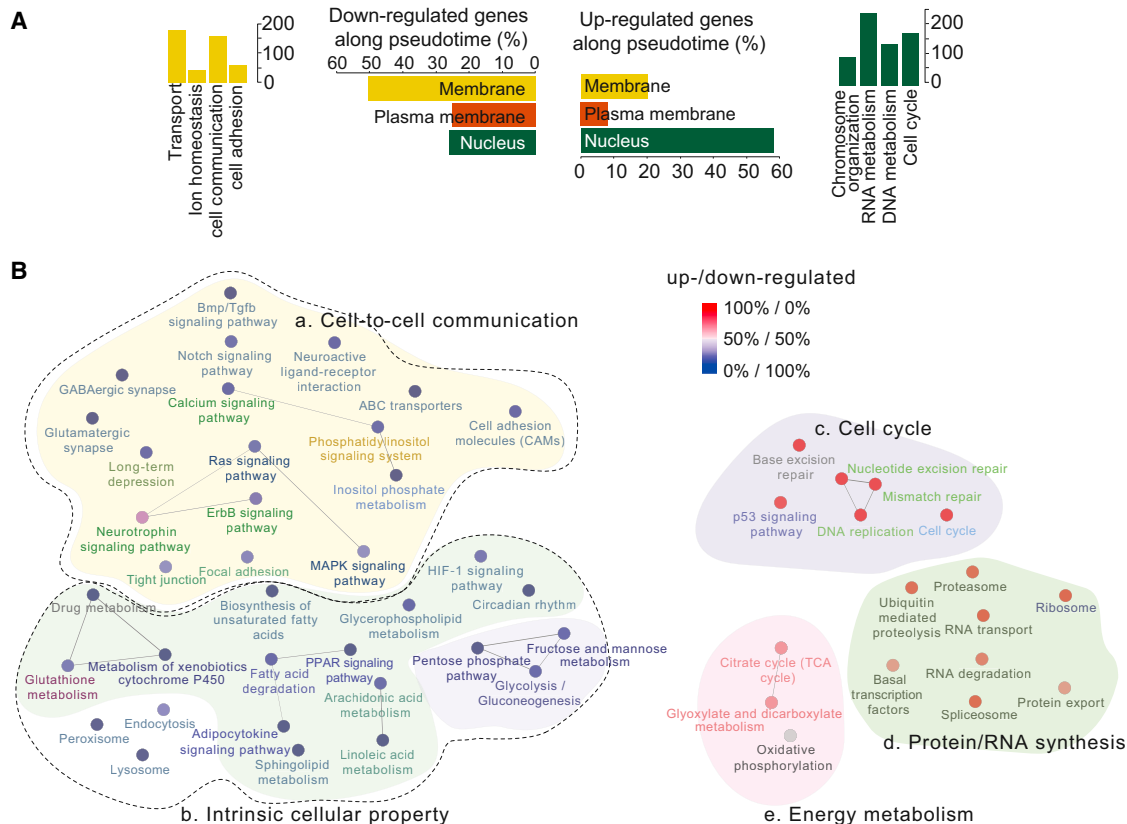
(legend on next page)

A

B

Figure 5. Functional Characterization of UP[1000] Genes and DOWN[1000] Genes

(A) Quantification of predicted cellular location of gene products of UP[1000] genes and DOWN[1000] genes (two middle panels). Also shown are numbers of genes with indicated functions for membrane-associated DOWN genes (left panel) and those for nucleus-associated UP genes (right panel).

(B) Functional GO analysis for DOWN[1000] and UP[1000] genes. The color of each functional entity represents the proportion of UP genes (blue) and DOWN genes (red). Connections between each pair of data points represent sharing more than five genes between the pair. Functionally similar entities are grouped with same background colors. Broken lines represent two categories of DOWN genes: genes encoding proteins involved in the intra- or extra-cellular communication and genes encoding proteins defining intrinsic stem cell properties.

See also Figures S5, S6, and S7.

resources. First, we provide a comprehensive dataset of single-cell transcriptomes of qNSCs and their immediate progeny in the adult mouse hippocampus in vivo. Second, we provide Waterfall, a bioinformatic pipeline for in silico reconstruction of molecular trajectories based on snapshots of single-cell transcriptomes and statistical gene expression analysis over continuous developmental processes. Third, we provide a holistic picture of adult qNSC molecular signatures and dynamic molecular cascades underlying initial phases of adult neurogenesis at unprecedented temporal resolution (Figure 7). Our study provides an example of how to resolve cellular heterogeneity and reveal developmental dynamics for systematic molecular characterization of stem cells and their differentiation in vivo. Our approach can be adapted for various single-cell omics analyses (transcriptomics, proteomics, epigenomics, lipidomics, and metabolo-

mics) of many continuous biological processes, such as development, physiological and pharmacological stimulation, and disease progression (see examples in Supplemental Experimental Procedures).

A Resource of In Vivo Single-Cell Transcriptomes of qNSCs and Their Immediate Progeny

Our study provides a single-cell RNA-seq dataset and the first comprehensive view of transcriptome dynamics underlying adult qNSC behavior in vivo. Currently, there is no published dataset for transcriptome dynamics during stem cell development in any somatic system in vivo. We performed multiple levels of validation of our dataset and approach, including comparison with an in situ database, confirmation with known and unknown marker expression during adult neurogenesis in vivo, and

Figure 4. Molecular Cascade underlying Adult qNSC Activation and Neurogenesis Initiation

(A) Lists of DOWN and UP TFs and their Spearman correlation coefficient with pseudotime.

(B) On/High (yellow) or Off/Low (black) states of top 150 DOWN (left) and UP (right) genes sorted by the timing of transition points. Shown on the top are histograms of the numbers of individual cells examined along the pseudotime progression. The colors on the histogram follow the color scheme of Figure 2B.
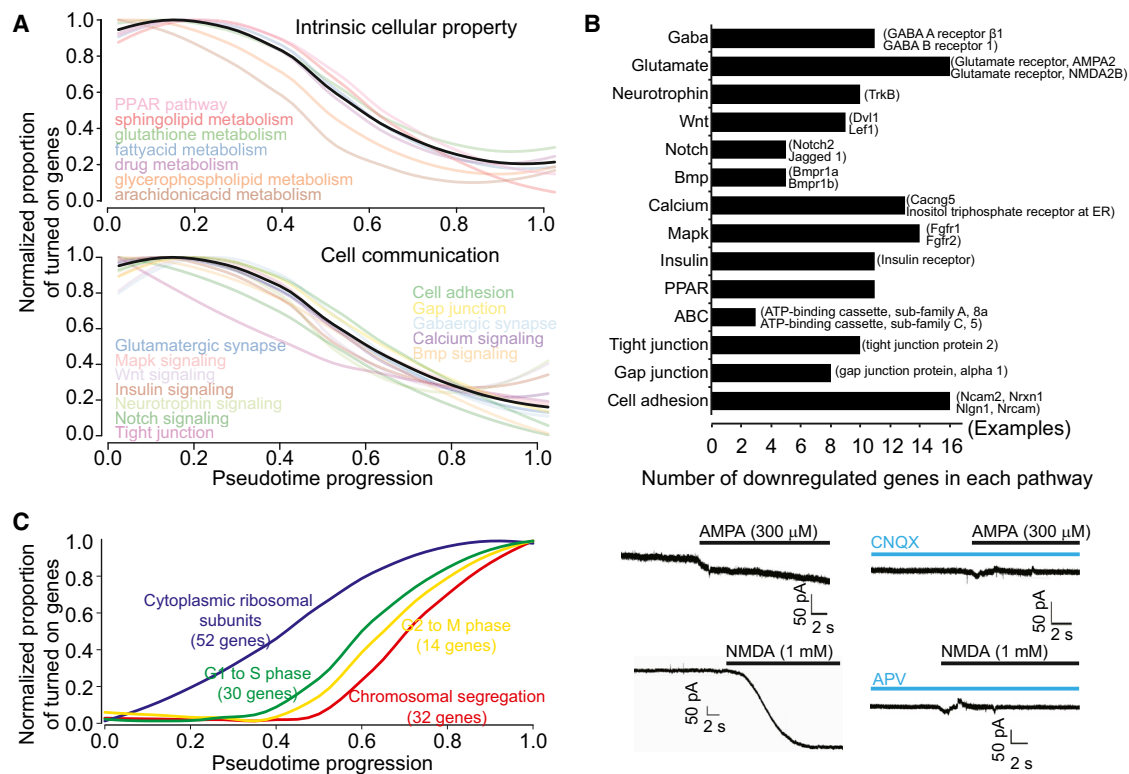
See also Figure S4.

Figure 6. Sequential Molecular Dynamics during Adult NSC Activation and Neurogenesis

(A) Gradual downregulation of averaged binary states of each functional entity defining intrinsic stem cell properties (top) and those defining intra- or extra-cellular communication (bottom). On/off binary states of expressed genes within each functional entity were averaged and normalized to show the timing of the transition.

(B) Number of genes in different DOWN GO groups and representative example genes. Expression patterns of representative genes over developmental pseudotime are shown in Figure S6A. Electrophysiological recording of Nestin-GFP^{cyto+} NSCs in acute hippocampal slices showed responses to both AMPA and NMDA (bottom).

(C) Gradual upregulation of averaged binary states of each functional entity defining cell-cycle checkpoints and cytoplasmic ribosomal subunits. On/off binary states of upregulated genes within each functional entity are averaged and normalized to exemplify the timing of the transition. Cell-cycle checkpoint genes were upregulated in the sequence of the cell-cycle progression. The upregulation of cytoplasmic ribosomal subunits preceded upregulation of the earliest cell-cycle checkpoint gene.

See also Figure S6.

functional validation via clonal lineage tracing and electrophysiology (Figures 3 and 6B; Table S3).

Our resource of holistic molecular profiles during early neurogenesis has three unique features that increase its versatility. First, our whole-transcriptome information includes unannotated transcripts, isoforms, and retrotransposon-derived transcripts, as opposed to multiplexed qPCR or microarray-based studies, which can only provide limited annotated transcripts (Hoppe et al., 2014). Second, we animated static single-cell transcriptomes over the in vivo neurogenesis trajectory, allowing queries of molecular dynamics for each gene over the continuous biological process and generation of novel hypotheses during specific phases of adult qNSC maintenance, activation, and neurogenesis initiation. For example, expression of nuclear glucocorticoid receptor Nr3c1 in adult qNSCs, but not in eIPCs (Figure 4A), suggests a cellular target of glucocorticoids during adult hippocampal neurogenesis and a means to manipulate qNSCs in vivo. Our resource also reveals potential prospective markers of adult neurogenesis, such as *Hopx-CreER*^{T2}, which can target adult qNSCs in vivo (Figures 3E–3G). Third, each transcriptome in

our dataset represents a biological state within a single cell. This modular construction allows for flexible reorganization of the dataset to probe different questions as our understanding of the biology evolves. For example, instead of generating reporter lines for individual genes or using different surface markers for physical sorting of specific cell populations, investigators can perform unlimited in silico cell sorting with any individual gene, or multiple genes in combination, to obtain a selected cell population to be probed for their gene expression characteristics at the genome-wide level using our single-cell datasets.

Developmental Dynamics of Adult Neurogenesis at the System Level

Recent genome-wide studies have begun to provide a system-level understanding of in vivo adult NSC biology using marker-defined NPC populations (Bracko et al., 2012; Codega et al., 2014; Kriegstein and Alvarez-Buylla, 2009). Yet previous studies have only acquired snapshots of transcriptomes, which limits investigation of developmental dynamics among different
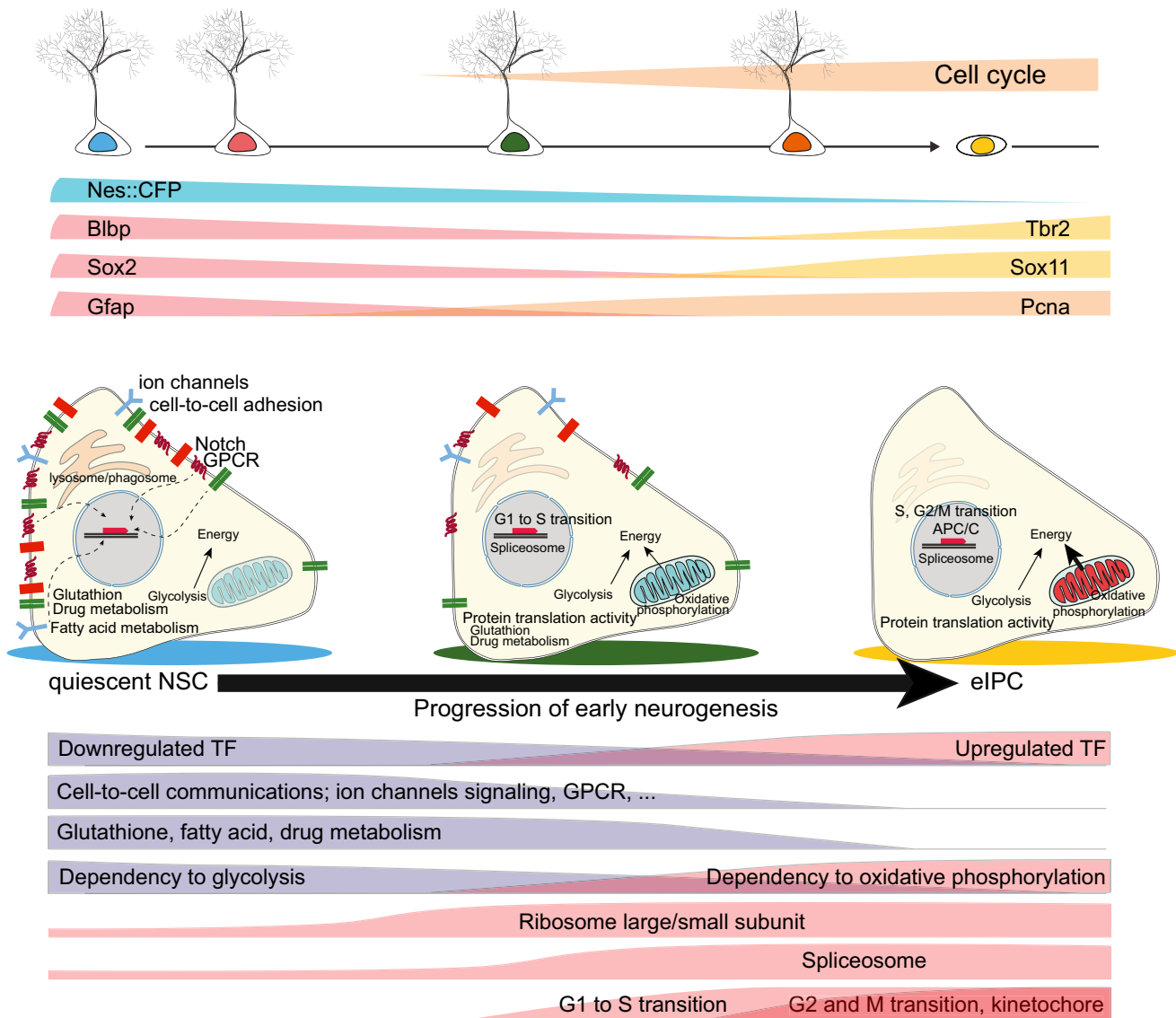
Cell cycle

Nes::CFP

Blbp — Tbr2

Sox2 — Sox11

Gfap — Pcna

ion channels
cell-to-cell adhesion
Notch
GPCR
lysosome/phagosome
Energy
Glutathion
Drug metabolism
Glycolysis
Fatty acid metabolism

quiescent NSC

G1 to S transition
Spliceosome
Energy
Glycolysis
Oxidative phosphorylation
Protein translation activity
Glutathion
Drug metabolism

S, G2/M transition
APC/C
Spliceosome
Energy
Glycolysis
Oxidative phosphorylation
Protein translation activity

eIPC

Progression of early neurogenesis

Downregulated TF — Upregulated TF

Cell-to-cell communications; ion channels signaling, GPCR, ...

Glutathione, fatty acid, drug metabolism

Dependency to glycolysis — Dependency to oxidative phosphorylation

Ribosome large/small subunit

Spliceosome

G1 to S transition — G2 and M transition, kinetochore

Figure 7. Schematic Summary of Molecular Signatures of Adult qNSCs and Molecular Cascades underlying Their Activation and Neurogenesis

Shown on top is an illustration of known marker expression and cell-cycle activation during adult hippocampal neurogenesis. Shown in the middle is an illustration of molecular signatures of adult qNSCs and their immediate progeny. Shown at the bottom are functional categories of genes that show a clear shift during adult qNSC activation and generation of eIPCs. qNSCs exhibit intra- and inter-cellular signaling to actively sense the local niche, rely mostly on glycolysis for energy, and have highly active fatty acid, glutathione, and drug metabolism. Upon activation, NSCs increase translational capacity, followed by cell-cycle entry with G1 to S transition. Oxidative phosphorylation starts to be active and stem-cell-specific properties are downregulated. eIPCs maintain active cell-cycle genes, ribosomal activity, and fully active oxidative phosphorylation for energy generation. The color scheme on the top and the middle illustration follows that of Figure 2B.

cellular states. We co-opted the imperfection of the *Nestin-CFP^nuc* genetic labeling system to collect individual qNSCs and their immediate neuronal progeny concurrently. Aligning cells along the developmental trajectory yielded, for the first time, a molecular continuum with sequential progression of the individual transcriptome from qNSC to aNSC and then eIPC. Importantly, this novel approach does not divide developmental processes into discrete stages that are defined a priori by capturing populations sharing specific markers.

Our resources provide unparalleled temporal resolution to identify new mechanisms underlying adult NSC biology. For

example, we showed that Acyl-CoA synthetases (Acsl3, Acsl6, and Acsbg1), the enzymes for the first step of fatty acid β-oxidation, were highly expressed only in qNSCs (Figure S6B), suggesting a novel role for active fatty acid β-oxidation in qNSCs and thereby extending previous findings on the role of fatty acid metabolism in adult neurogenesis (Knobloch et al., 2013) (Table S2). We also found that ribosomal subunits were the first genes upregulated upon adult qNSC activation and early differentiation (Figures 6C and S6D), suggesting a possible demarcation of G0 to G1 transition and providing the timing for switches in protein synthesis, the regulation of which is important for

somatic stem cell function (Signer et al., 2014). The holistic picture we obtained unifies disparate information and illuminates novel biological themes in stem cell biology. For example, our analysis suggests that qNSCs actively respond to local environmental cues through various signaling pathways, but gradually and globally shut off signaling capacity upon activation. These observations support the concept of a niche wherein mammalian somatic stem cells are tightly controlled by a regulatory microenvironment (Schofield, 1978) and predict that eIPCs are less responsive to environmental input (Berg et al., 2015). This novel biological insight may be applicable to many somatic stem systems defined by stochastic behavior (Simons and Clevers, 2011).

Waterfall Analysis of Single-Cell Transcriptomes within a Continuum

Waterfall has three key differences from previous methodologies. First, it requires very little prior information to generate a highly accurate temporal trajectory at single-cell resolution. Previous methods have been able to reconstruct accurate trajectories by relying on a robust set of known markers to establish cell order and validate cell alignment at numerous points along the timeline. For many biological systems and processes, we have much less information. Second, in contrast to Monocle (Trapnell et al., 2014) or Wanderlust (Bendall et al., 2014), Waterfall uses k-means clustering to build a trajectory and assign an individual cell a pseudotime based on each cell's proximity to the cluster-derived trajectory, rather than construct a trajectory by directly connecting each cell to the next. Third, in order to analyze stochastic gene expressions, we adopted HMM to predict consecutive binary states in gene expression activity over pseudotime. HMM permits the interpretation of highly variable data without logarithmic transformation, normalization, or the input of any arbitrary parameters, such as threshold for gene expression noise or Markovian parameters (transition probability, initial probability, and emission probability). Our validation for known and unknown genes in adult neurogenesis indicated that HMM correctly predicted temporal dynamics of in vivo biology.

There is no conceptual restriction of our approach to transcriptome studies of adult neurogenesis. Indeed, in the Supplemental Experimental Procedures we provide examples of how Waterfall could be broadly applicable for single-cell RNA-seq datasets from analyses of in vitro myogenesis and in vivo embryonic lung development, for a single-cell mass-cytometry dataset from analysis of in vivo B cell development, and for two synthetic datasets. We expect that Waterfall algorithms can be adopted for diverse single-cell multi-dimensional datasets, including single-cell transcriptomes, epigenomes, proteomes, and metabolomes, of various continuous biological processes.

EXPERIMENTAL PROCEDURES

Preparation of Individual Cells from Adult Mouse Dentate Gyrus

Homozygous transgenic mice expressing nuclear localized CFP (CFP^{nuc}) driven by the Nestin regulatory elements (Encinas et al., 2006) were used for all single-cell RNA-seq experiments. Mice were euthanized by cervical dislocation, and brains were immediately immersed into ice cold Dulbecco's Phosphate-Buffered Saline (DPBS, Corning). All procedures were performed with approved protocols in accordance with institutional animal guidelines.

The dissected dentate gyrus was incubated in Hibernate A (BrainBits) containing papain (100 U; Sigma) and RNase-free DNase I (100 units; NEB) at 37°C for 18 min with intermittent flicking. The tissue was triturated into individual cell suspension by 1 ml pipette (Denville Scientific). Enzymes and cellular debris were removed with multiple rounds (~4–5 times) of mild centrifugation at 200 g and washing with Hibernate A minus Ca^{2+} and Mg^{2+} (BrainBits). The individual cell suspension was plated onto a glass bottom plate (MatTek) and picked up using glass pipettes (World Precision Instruments) under a fluorescent microscope. The glass tip was broken into the bottom of each PCR tube containing water (2.4 μl) with RNase-free DNase I (0.2 μl; NEB) and Murine origin RNase inhibitor (0.25 μl; NEB). Importantly, the addition of DNase I significantly improved the quality of the data by removing contamination from the random amplification of genomic DNA (see Supplemental Experimental Procedures).

Library Preparation and Sequencing

cDNA amplification followed the previously published SMART protocol (Ramsköld et al., 2012). Briefly, we first inactivated the DNase I by increasing the temperature (75°C for 10 min), and samples were then stored on ice. Custom designed 2A oligo 1 μl (12 μM, Integrated DNA Technologies, sequence shown in Figure S1A) was added and annealed to the polyadenylated RNA by undergoing a temperature increase (75°C for 3 min) and being quenched on ice. A mixture of 2 μl Superscript II First-Strand Buffer (5×, Invitrogen), 1 μl custom designed TS oligo (12 μM, Integrated DNA Technologies, Figure S1A), 0.3 μl MgCl$_2$ (200 mM, Sigma), 0.5 μl RNase inhibitor (Neb), 1 μl dNTP (10 mM each, Thermo), 0.25 μl DTT (100 mM, Invitrogen), and 1 μl Superscript II (200 U/μl, Invitrogen) were incubated at 42°C for 90 min, which was followed by enzyme inactivation at 75°C for 10 min. A mixture of 29 μl water, 5 μl Advantage2 taq polymerase buffer, 2 μl dNTP (10 mM each, Thermo), 2 μl custom-designed PCR primer (12 μM, Integrated DNA Technologies, Figure S1A), and 2 μl Advantage2 taq polymerase was directly added to the reverse transcription product and the amplification was performed for 19 cycles. The amplification product was purified using Ampure XP beads (Beckman-Coulter). Library preparation was performed using Ovation Ultralow library systems (Nugen Inc.). Libraries were multiplexed and sequenced using Illumina Hiseq 2500 (Illumina Inc) (Table S1).

Bioinformatic Analyses
Mapping and Calculating Gene Expression Levels
Raw reads were trimmed for the Illumina adaptor sequences using Trimmomatic (Bolger et al., 2014) and trimmed for the 5′ TS oligo sequences and the 3′ primer sequences using custom R codes. RSEM (Li and Dewey, 2011) was used to map and calculate gene expression levels represented as transcripts per million (TPM). The reference genome was modified to include chrC, which contained the sequence of part of the Nestin enhancer followed by eCFP transcripts, and reconstructed from sequencing reads. We used the following parameters: rsem-calculate-expression -p 12–fragment-length-mean 500 $input.fastq $rsem_ref $cell_id. For downstream analyses, we used a table with single cells at the columns and the genes at the rows (Table S6).
Waterfall 1: Pre-Processing
Waterfall input is an expression matrix from RSEM after the elimination of outliers (Figure S1C). Unsupervised clustering was performed using a distance matrix based on Pearson correlation between each pair of single cells (Figure 2A). We defined the neurogenic trajectory on the PCA plot and determined the direction using known markers such as Sox11, Tbr2, Blbp, and Gfap.
Waterfall 2: Building an In Vivo Trajectory
We used custom R codes (see file Data S1) to determine pseudotime for each single cell on the trajectory (Figure S3A; also see Supplemental Experimental Procedures). Briefly, we performed parametric PCA and extracted k-means from the distribution of single-cell transcriptomes. We generate an unbiased trajectory by connecting k-means centers using an MST algorithm (Paradis et al., 2004). First, we set zero for the origin of the continuous trajectory, determined by pre-processing. Second, we assigned locations to individual cell data points on the trajectory. We assigned each cell to the closest MST segment (lines between k-means) or vertex (k-mean) with a single perpendicular projection. Third, we straightened all the segments into one horizontal line and determined the relative order of the assigned locations of single cell data

points on the trajectory. Pseudotime values ranged from 0 (at the origin) to 1 (at the end).

Waterfall 3: Gene Expression Analysis by HMM

We used custom R codes to apply an HMM to predict gene expression states throughout pseudotime. Briefly, we divided pseudotime into 40 bins, each of which contained an average of 2.5 single cells. We averaged the expression level within each bin and assigned the expression values to observed variables for HMM. We used a Baum-Welch algorithm to extract the most probable emission probabilities and transition probabilities. Using the output from the Baum-Welch algorithm along with observed variables, we applied the Viterbi algorithm to predict binary gene expression states (Figure S3B).

Functional Gene Expression Analysis

We calculated the Spearman correlation coefficient between pseudotime points and each gene's expression TPM values. Genes with relatively high Spearman correlations were defined as UP genes and genes with relatively low correlations were DOWN genes, and the highest and lowest 1,000 genes were defined as UP^{1000} and DOWN^{1000}, respectively (Table S4). A small subset of the UP^{1000} and DOWN^{1000} genes with low average expression values (<50 TPM) and low coefficient of variation (<1.95) were from repeat elements within their exons and were excluded from downstream analyses. Raw mapping profiles of all genes shown in pseudotime figures were closely inspected to rule out false positives. We identified TFs using public databases (Zhang et al., 2012). We used GO (Ashburner et al., 2000), KEGG (Kanehisa and Goto, 2000), and wikipathway (Pico et al., 2008) for functional GO analyses, along with the R bioconductor package (Gentleman et al., 2004) or Cytoscape software (Shannon et al., 2003).

For alternative functional gene expression analysis (Figure S7), we first divided the entire transcriptome dataset into three equal groups based on their Spearman correlation to pseudotime: positively correlated, uncorrelated, and negatively correlated (Table S7). We then evaluated the proportion of positively correlated genes versus negatively correlated genes within each functional entity from an independent functional annotation database wikipathway rather than the KEGG pathway database. If a functional entity contained a disproportionally larger number of upregulated genes than downregulated genes, we considered the functional entity to be generally activated, and, conversely, a disproportionally larger number of downregulated genes indicated that the pathway was generally inactivated over time.

Validation with In Situ Database Comparison, Immunohistology, Genetic Labeling, and Electrophysiology

We validated our NPC-enriched genes by the Allen mouse brain atlas in situ hybridization dataset (Lein et al., 2007) (Table S3). We inspected the gene expression patterns within the adult dentate gyrus at sagittal views. Genes with clear and relatively even distribution within the SGZ were determined to be "SGZ enriched," whereas genes with subtle or scattered enrichment within the SGZ were determined to be "ambiguous." Genes without any enrichment at the SGZ were determined to be "not SGZ enriched."

Adult Nestin-GFP^{cyto} animals were used for immunohistochemical validation. Hopx-CreER^{T2} $^{f/+}$;;mT/mG$^{f/+}$ mice were generated by crossing Hopx-CreER^{T2} $^{f/+}$ (Takeda et al., 2011) [strain: Hopx$^{tm2.1(cre/ERT2)Joe}$/J, Jackson Labs Stock: 017606] with the mT/mG$^{f/f}$ reporter line [strain: B6.129(Cg)-Gt(ROSA)26Sor$^{tm4(ACTB-tdTomato,-EGFP)Luo}$/J; Jackson Labs Stock: 007676]. Tamoxifen (62 mg/ml; Sigma; T5648) was prepared in a 5:1 ratio of corn oil/ethanol, heated to 37°C, and mixed. Eight-week-old HopX-CreER^{T2} $^{f/+}$;;mT/mG$^{f/+}$ animals were injected intraperitoneally with 124 mg/kg tamoxifen and analyzed by immunohistology as previously described (Bonaguidi et al., 2011). The following antibodies were used: Aldoc (1:200, goat; Cat#SC12065; Santa Cruz), GFAP (1:2,000, rabbit; Cat#Z0334; DAKO), GFP (1:1,000, chicken; Cat#GFP-1020; Aves), GFP (1:1,000, goat; Cat#600-101-215; Rockland), Nestin (1:500, chicken; Cat#NES; Aves), PCNA (1:2,000, rabbit; Cat#ab18197; Abcam), PCNA (1:500, goat; Cat#SC9857; Santa Cruz), Stmn1 (1:200, rabbit; Cat#ab24445; Abcam), and Tbr2 (1:1,000, rabbit; Cat#Ab23345; Abcam). GFP cells were identified with an Axiovert 200M microscope (Zeiss) and then acquired as z-stacks on Zeiss 710 single-photon confocal microscope using 40× or 63× objectives. For quantification of Stmn1, Aldoc, and PCNA expression in Nestin-GFP^{cyto} mice, z-stacks were acquired from three animals. Images were analyzed using Imaris 7.1.1 (Bitplane). Radial-glia-like NSCs were identified by their radial processes and somas situated in the SGZ and IPCs were identified by their small somas and tangential processes as previously described (Bonaguidi et al., 2011).

Adult Nestin-GFP^{cyto} transgenic mice were used to validate expression of functional glutamate receptors on NSCs. GFP$^+$ radial-glia-like NSCs in slices prepared acutely from adult animals were recorded by whole-cell patch-clamp upon puffing of AMPA or NMDA in the presence or absence of antagonists as previously described (Song et al., 2012).

REFERENCES

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al.; The Gene Ontology Consortium (2000). Gene ontology: tool for the unification of biology. Nat. Genet. 25, 25–29.

Bendall, S.C., Davis, K.L., Amir, A.D., Tadmor, M.D., Simonds, E.F., Chen, T.J., Shenfeld, D.K., Nolan, G.P., and Pe'er, D. (2014). Single-cell trajectory detection uncovers progression and regulatory coordination in human B cell development. Cell 157, 714–725.

Berg, D.A., Yoon, K.-J., Will, B., Xiao, A.Y., Kim, N.-S., Christian, K.M., Song, H., and Ming, G.-l. (2015). Tbr2-expressing intermediate progenitor cells in the adult mouse hippocampus are unipotent neuronal precursors with limited amplification capacity under homeostasis. Frontiers in Biology 10, 262–271.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120.

Bonaguidi, M.A., Wheeler, M.A., Shapiro, J.S., Stadel, R.P., Sun, G.J., Ming, G.L., and Song, H. (2011). In vivo clonal analysis reveals self-renewing and multipotent adult neural stem cell characteristics. Cell 145, 1142–1155.

Bracko, O., Singer, T., Aigner, S., Knobloch, M., Winner, B., Ray, J., Clemenson, G.D., Jr., Suh, H., Couillard-Despres, S., Aigner, L., et al. (2012). Gene expression profiling of neural stem cells and their neuronal progeny reveals IGF2 as a regulator of adult hippocampal neurogenesis. J. Neurosci. 32, 3376–3387.

Codega, P., Silva-Vargas, V., Paul, A., Maldonado-Soto, A.R., Deleo, A.M., Pastrana, E., and Doetsch, F. (2014). Prospective identification and purification of quiescent adult neural stem cells from their in vivo niche. Neuron 82, 545–559.

Encinas, J.M., Vaahtokari, A., and Enikolopov, G. (2006). Fluoxetine targets early progenitor cells in the adult brain. Proc. Natl. Acad. Sci. USA 103, 8233–8238.

Encinas, J.M., Michurina, T.V., Peunova, N., Park, J.H., Tordo, J., Peterson, D.A., Fishell, G., Koulakov, A., and Enikolopov, G. (2011). Division-coupled astrocytic differentiation and age-related depletion of neural stem cells in the adult hippocampus. Cell Stem Cell *8*, 566–579.

Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., et al. (2004). Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. *5*, R80.

Hoppe, P.S., Coutu, D.L., and Schroeder, T. (2014). Single-cell technologies sharpen up mammalian stem cell research. Nat. Cell Biol. *16*, 919–927.

Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. *28*, 27–30.

Knobloch, M., Braun, S.M., Zurkirchen, L., von Schoultz, C., Zamboni, N., Araúzo-Bravo, M.J., Kovacs, W.J., Karalay, O., Suter, U., Machado, R.A., et al. (2013). Metabolic control of adult neural stem cell activity by Fasn-dependent lipogenesis. Nature *493*, 226–230.

Kriegstein, A., and Alvarez-Buylla, A. (2009). The glial nature of embryonic and adult neural stem cells. Annu. Rev. Neurosci. *32*, 149–184.

Lein, E.S., Hawrylycz, M.J., Ao, N., Ayres, M., Bensinger, A., Bernard, A., Boe, A.F., Boguski, M.S., Brockway, K.S., Byrnes, E.J., et al. (2007). Genome-wide atlas of gene expression in the adult mouse brain. Nature *445*, 168–176.

Li, L., and Clevers, H. (2010). Coexistence of quiescent and active adult stem cells in mammals. Science *327*, 542–545.

Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics *12*, 323.

Lu, R., Neff, N.F., Quake, S.R., and Weissman, I.L. (2011). Tracking single hematopoietic stem cells in vivo using high-throughput sequencing in conjunction with viral genetic barcoding. Nat. Biotechnol. *29*, 928–933.

Mignone, J.L., Kukekov, V., Chiang, A.S., Steindler, D., and Enikolopov, G. (2004). Neural stem and progenitor cells in nestin-GFP transgenic mice. J. Comp. Neurol. *469*, 311–324.

Ming, G.L., and Song, H. (2011). Adult neurogenesis in the mammalian brain: significant answers and significant questions. Neuron *70*, 687–702.

Muramoto, T., Cannon, D., Gierlinski, M., Corrigan, A., Barton, G.J., and Chubb, J.R. (2012). Live imaging of nascent RNA dynamics reveals distinct types of transcriptional pulse regulation. Proc. Natl. Acad. Sci. USA *109*, 7350–7355.

Novick, A., and Weiner, M. (1957). Enzyme Induction as an All-or-None Phenomenon. Proc. Natl. Acad. Sci. USA *43*, 553–566.

Paradis, E., Claude, J., and Strimmer, K. (2004). APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics *20*, 289–290.

Pico, A.R., Kelder, T., van Iersel, M.P., Hanspers, K., Conklin, B.R., and Evelo, C. (2008). WikiPathways: pathway editing for the people. PLoS Biol. *6*, e184.

Raj, A., Peskin, C.S., Tranchina, D., Vargas, D.Y., and Tyagi, S. (2006). Stochastic mRNA synthesis in mammalian cells. PLoS Biol. *4*, e309.

Ramsköld, D., Luo, S., Wang, Y.C., Li, R., Deng, Q., Faridani, O.R., Daniels, G.A., Khrebtukova, I., Loring, J.F., Laurent, L.C., et al. (2012). Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. Nat. Biotechnol. *30*, 777–782.

Schofield, R. (1978). The relationship between the spleen colony-forming cell and the haemopoietic stem cell. Blood Cells *4*, 7–25.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. *13*, 2498–2504.

Signer, R.A., Magee, J.A., Salic, A., and Morrison, S.J. (2014). Haematopoietic stem cells require a highly regulated protein synthesis rate. Nature *509*, 49–54.

Simons, B.D., and Clevers, H. (2011). Strategies for homeostatic stem cell self-renewal in adult tissues. Cell *145*, 851–862.

Song, J., Zhong, C., Bonaguidi, M.A., Sun, G.J., Hsu, D., Gu, Y., Meletis, K., Huang, Z.J., Ge, S., Enikolopov, G., et al. (2012). Neuronal circuitry mechanism regulating adult quiescent neural stem-cell fate decision. Nature *489*, 150–154.

Stegle, O., Teichmann, S.A., and Marioni, J.C. (2015). Computational and analytical challenges in single-cell transcriptomics. Nat. Rev. Genet. *16*, 133–145.

Takeda, N., Jain, R., LeBoeuf, M.R., Wang, Q., Lu, M.M., and Epstein, J.A. (2011). Interconversion between intestinal stem cell populations in distinct niches. Science *334*, 1420–1424.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. *28*, 511–515.

Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N.J., Livak, K.J., Mikkelsen, T.S., and Rinn, J.L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat. Biotechnol. *32*, 381–386.

Zhang, H.M., Chen, H., Liu, W., Liu, H., Gong, J., Wang, H., and Guo, A.Y. (2012). AnimalTFDB: a comprehensive animal transcription factor database. Nucleic Acids Res. *40*, D144–D149.