SCIENTIFIC REPORTS

OPEN

SUBJECT AREAS:
EPIGENOMICS
NEURAL STEM CELLS

Received
14 February 2014

Accepted
2 June 2014

Published
20 June 2014

Correspondence and requests for materials should be addressed to C.-H.A.L. (annie.lin@utsa.edu)

Epigenetic Regulation by Chromatin Activation Mark H3K4me3 in Primate Progenitor Cells within Adult Neurogenic Niche

Richard S. Sandstrom[1], Michael R. Foret[2], Douglas A. Grow[2,3], Eric Haugen[1], Christopher T. Rhodes[2], Astrid E. Cardona[2], Clyde F. Phelix[2], Yufeng Wang[2], Mitchel S. Berger[4] & Chin-Hsing Annie Lin[2,3]

[1]Department of Genome Sciences, University of Washington, 1705 NE Pacific Street, Seattle, WA 98195, USA, [2]Department of Biology, University of Texas at San Antonio, One UTSA Circle, San Antonio, Texas 78249, USA, [3]Neuroscience Institute, University of Texas at San Antonio, San Antonio, Texas 78249, USA, [4]Department of Neurological Surgery, University of California at San Francisco, San Francisco, California 94143, USA.

Histone 3 lysine 4 trimethylation (H3K4me3) is known to be associated with transcriptionally active or poised genes and required for postnatal neurogenesis within the subventricular zone (SVZ) in the rodent model. Previous comparisons have shown significant correlation between baboon (*Papio anubis*) and human brain. In this study, we demonstrate that chromatin activation mark H3K4me3 is present in undifferentiated progenitor cells within the SVZ of adult baboon brain. To identify the targets and regulatory role of H3K4me3 within the baboon SVZ, we developed a technique to purify undifferentiated SVZ cells while preserving the endogenous nature without introducing culture artifact to maintain the *in vivo* chromatin state for genome-wide studies (ChIP-Seq and RNA-Seq). Overall, H3K4me3 is significantly enriched for genes involved in cell cycle, metabolism, protein synthesis, signaling pathways, and cancer mechanisms. Additionally, we found elevated levels of H3K4me3 in the MRI-classified SVZ-associated Glioblastoma Multiforme (GBM), which has a transcriptional profile that reflects the H3K4me3 modifications in the undifferentiated progenitor cells of the baboon SVZ. Our findings highlight the importance of H3K4me3 in coordinating distinct networks and pathways for life-long neurogenesis, and suggest that subtypes of GBM could occur, at least in part, due to aberrant H3K4me3 epigenetic regulation.

Brain regions retaining neural stem cells (NSCs) for postnatal and adult neurogenesis include the subventricular zone (SVZ) on the walls of the lateral ventricles and the subgranular zone within the dentate gyrus of the hippocampus[1–4]. Within the SVZ, there is a population of slowly dividing NSCs with astrocyte-like morphology. These NSCs give rise to transit-amplifying cells, which subsequently give rise to immature neuroblasts. In the rodent, neuroblasts ultimately migrate through the rostral migratory stream to generate interneurons in the olfactory bulb[5–7]. Numerous studies have demonstrated that extracellular signals[5,8–11], intracellular regulators[6,12,13], and epigenetic mechanisms[1] have significant effects on the self-renewal and lineage commitment of NSCs. An epigenetic mechanism through MLL1 (mixed-lineage leukaemia 1, a histone methyltransferase responsible for histone 3 lysine 4 trimethylation-H3K4me3) has been shown to induce postnatal neurogenesis in SVZ[14]. The MLL family methyltransferases (MLL1, MLL2, MLL3, MLL4, MLL5) reside within the Trithorax (Trx) group protein complex, which also contains WDR5/RbBP5/ASH2 subunits to convert the demethylated forms of H3K4 to the trimethylated form in order to render open chromatin[15,16]. To date, the emerging consensus from genome-wide scale analysis reveals that tri-methylation at H3K4 (H3K4me3) is characteristic of transcriptionally active or poised status[17,18]. Importantly, the H3K4me3/MLL family is also known to be associated with distinct types of hematologic disorders[19,20]. Thus, oncogenesis under MLL deregulation could result in changes of gene expression through the alteration of H3K4me3 epigenetic landscape.

In this work, we demonstrate that H3K4me3 is present in progenitor cells residing in the primate SVZ, where adult neurogenesis occurs. To uncover the function of H3K4me3 in this adult germinal niche, we have chosen to utilize the nonhuman primate- baboon (*Papio anubis*) for genome-wide ChIP-Seq (chromatin immunoprecipi-

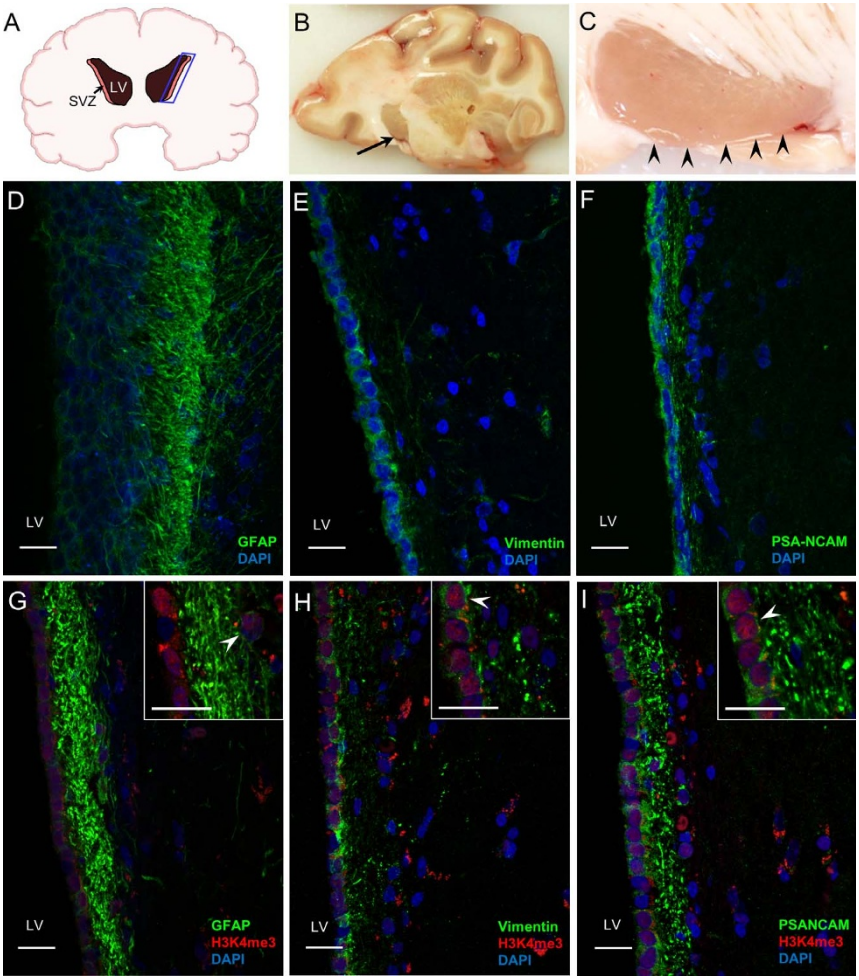

Figure 1 | Gross anatomy of baboon SVZ architecture and the colocalization of H3K4me3 and cell type specific markers in baboon SVZ. (A) Scheme displays coronal view of baboon brain with dictation of lateral ventricle (LV) and subventricular zone (SVZ). (B) Coronal slice of baboon forebrain showing regions sampled, the arrow points to the head of the caudate nucleus at the margin with the lateral ventricle. The dorsal is to the left and medial is at the bottom for orientation. 2× magnification. (C) Horizontal slice through the caudate nucleus showing the subventricular zone (SVZ) in temporal horn of lateral ventricle; arrow heads point to SVZ; 6× magnification. (D) Images of the caudal ventral subventricular zone reveal the GFAP-positive astrocytic ribbon. A deeper z-stack imaging was performed in order to view GFAP-positive processes of the NSCs extending toward the lateral ventricle (LV). (E) Vimentin staining shows NSC population in the baboon SVZ. (F) PSA-NCAM staining shows neuroblast population in the baboon SVZ. (G) A small population of H3K4me3 positive cells is colocalized with GFAP in rostral ventral SVZ. (H) Significant population of Vimentin-positive cells is colocalized with H3K4me3. (I) H3K4me3 persists in the PSA-NCAM positive neuroblast population. (D–I) 40× images; inserts are 100× magnification; White arrow heads in G and H indicate colocalization of H3K4me3 and cell type markers; (G–I) Images and inserts are 2um single slices. Scale bars present 20 μm.

tation and deep sequencing) and RNA-Seq analyses. We developed a technique to purify progenitor cell populations within the SVZ directly from baboon brain, in which the large brain mass allows us to isolate sufficient SVZ cells without culture artifacts. We identified that H3K4me3 is significantly enriched for genes involved in cell cycle, cell signaling, nervous system development, metabolism, and ribosomal biogenesis. Among these genes, RNA-Seq identified that genes associated with cellular signaling/maintenance, DNA replication, metabolism, and protein synthesis are expressed. Our findings highlight that H3K4me3 is critical for SVZ cellular activity of life-long neurogenesis. Importantly, we found that subtypes of SVZ-associated Glioblastoma Multiforme (GBM) have increased levels of H3K4me3, in which changes in subsets of gene expression under aberrant H3K4me3 activation were also identified in these two subtypes of GBM (group I and II GBM). We anticipate that aberrant regulation by H3K4me3 in this adult germinal niche may potentiate tumorigenesis in the context of neurogenesis as this supports the growing conviction of epigenetic alterations in malignancy[21,22].

Results

Chromatin activation mark H3K4me3 in primate progenitor cells within adult SVZ. We utilized the previously characterized NSCs and neuroblasts markers[23] including GFAP, Vimentin, and PSA-NCAM for gross anatomical examination of the SVZ in non-human primate-baboon (Fig. 1). We found that GFAP-positive astrocytic ribbon lines this niche and some GFAP-positive cells extend toward the ventricle (Fig. 1D), resembling the previous finding in adult human brain[23–25]. Both Vimentin (NSC marker) and PSA-NCAM (neuroblast marker) also label the SVZ cells in the baboon brain (Fig. 1E, F). To examine which cell population in the baboon SVZ is enriched with H3K4me3, we applied co-immunostaining for H3K4me3 and cell type specific markers- GFAP, Vimentin, and PSA-NCAM representing NSCs and neuroblasts, respectively. We found that H3K4me3 is associated with small population of GFAP-positive NSCs, colocalized with Vimentin-positive cells, and persistent in PSA-NCAM-positive neuroblasts (Fig. 1G–I). Given the significant correlation of brain architecture and genomic level between baboon and human[26,27], we reason findings in baboon are relevant to human. In the parallel experi-

ment, we also found that H3K4me3 is present in the adult mouse SVZ co-localizing with GFAP, Mash1, and doublecortin (DCX)/PSA-NCAM representing NSCs, transit-amplifying progenitors, and neuroblasts, respectively (Supplementary Fig. 1). We performed a 2-hour EdU administration in mice to label quickly dividing transit-amplifying cells and neuroblasts, and found colocalization of EdU and H3K4me3 (Supplementary Fig. 1D). Further, we performed a cytosine-β-D-arabinofuranoside (AraC) ablation to validate the H3K4me3 presence in the SVZ. Antimitotic treatment via the infusion of AraC into adult mouse brain eliminates the quickly dividing transit amplifying cells and neuroblasts in the SVZ but leaves slowly dividing stem cells unaffected[28,29]. Notably, the population of H3K4me3 positive cells was significantly decreased (Supplementary Fig. 1), indicating that H3K4me3 is associated with transit amplifying cells and neuroblasts in addition to NSCs. Our results reveal that previously characterized markers for distinct SVZ cells in the rodent and human brain can also represent specific cell types in the baboon SVZ, and that H3K4me3 is present in the undifferentiated SVZ cells consistent across species.

H3K4me3 epigenetic landscape in primate progenitor cells within adult SVZ. Given the presence of H3K4me3 in the baboon SVZ cells, the molecular targets of this chromatin active mark in adult SVZ remain to be determined. To this end, we sought to identify the genomic loci that carry the H3K4me3 modification in the SVZ cells of the baboon, which share 92% sequence homology to humans at the genomic level. Additionally, comparisons across primate species and human have shown significant correlation in terms of brain volume and structure between baboon and human[26,27]. However, the spatial location and complexity of cell types in SVZ limit the dissection of discrete populations of cells within this niche for analysis. Nevertheless, the *in vitro* cell culture system cannot recapitulate *in vivo* cellular and molecular mechanisms that regulate identity and plasticity of the SVZ cells. To overcome this limitation and to ascertain characteristics of the SVZ cells as they exist *in vivo*, we developed a technique to purify SVZ cells directly from baboon brain within a short post-mortem interval (<20 minutes). After SVZ dissection, dissociated cells were analyzed by flow cytometry that showed ~8% GFAP-positive (NSC marker) and 15% DCX-positive (early and migrating neuroblast marker) without mature neuron contamination from nearby caudate (Fig. 2A). For purification, we manually conjugated Dynabeads to antibodies against SVZ cell type-specific markers, and then used the antibody-conjugated Dynabeads to purify dissociated cells following dissection of the SVZ. During developing this purification technique, we ran flow cytometry to validate the recovery of NSC and neuroblast cells. In the population purified by Vimentin-conjuagated Dynabeads, 59.8% of eluted cells expressed GFAP, and there was no DCX^+ cells (neuroblast marker) in this population. Likewise, 56.5% Vimentin-positive cells were recovered from GFAP-Dynabeads purification (Fig. 2C). These are not surprising results since GFAP labels both quiescent and active NSCs and Vimentin labels relatively active NSCs. As DCX labels both early and migrating neuroblasts and PSA-NCAM labels migrating neuroblasts, Dynabeads-DCX recovered 63% PSA-$NCAM^+$ population (Fig. 2D). To ensure no contamination of matured cells from the neighboring striatum, we also examine the presence of GAD65, which labels the mature neurons. We found no detectable level of GAD65 by flow cytometry. Our technical innovation preserves the *in vivo* nature of undifferentiated SVZ cells and is ideal for uncovering H3K4me3 enriched loci through the genome-wide ChIP-Seq approach. Each set of purified NSCs, neuroblasts, or combined progenitor cells (including NSCs and neuroblasts) from SVZ were subjected to ChIP with an antibody specific for H3K4me3 and sequenced to high depth (Fig. 3A). The baboon gene annotation is not currently available, however, there is only 2% difference at genomic level between baboon and rhesus macaque. Thus, the resulting reads were alternatively compiled, processed, and aligned to the Jan. 2006 rhesus macaque (Macaca mulatta) draft assembly, Mmul_051212[30] (http://www.ncbi.nlm.nih.gov/genome/assembly/237568/) and the UCSC genome browser version: rheMac2. Hotspot was used to identify H3K4me3-enriched loci (FDR = 0.01)[31]. Subsequently, H3K4me3-enriched genes were identified to be associated to their closest RefSeq gene and putative transcription start sites (TSS) by using the closest-features program from the BEDOPS tool set[32] (Fig. 3B–F; Supplementary Table 1). Based on rhesus macaque annotation, the enrichment of H3K4me3 located less than 1 kb away from annotated TSS is approximately 70% of H3K4me3 target genes, and ~8% of H3K4me3-enriched regions are >10 kb away from annotated TSS. Therefore, the distribution of H3K4me3 enrichment is preferentially close to TSS in the undifferentiated SVZ cells. Gene ontology (GO) analysis for the pool of total progenitor cells including NSCs and neuroblasts reveals that molecular functions of these genes include cell cycle, cell signaling, development, metabolism, neuronal differentiation, and ribosomal biogenesis (Fig. 3G and Supplementary Table 1). In addition, the top biological networks predicted by Ingenuity Pathways Analysis (IPA) are involved in 1) nervous system development and function; 2) post-transcriptional and post-translational modifications; 3) molecular transport; 4) ribosomal biogenesis; 5) metabolism (Fig. 3G and Supplementary Fig. 2–5). It is of interest to distinguish the differential regulation by H3K4me3 in NSC- versus neuroblast-specific populations. We therefore conducted 2 additional ChIP-Seq experiments for purified NSCs and neuroblasts separately. Importantly, the overrepresented genes in the purified NSC population are significantly enriched in categories of cell cycle, cell death, DNA repair, and proliferation, whereas the statistical over-representation among genes in the purified neuroblast population is highly associated with ribosomal biogenesis (Supplementary Fig. 6 and Supplementary Table 1). Among the shared targets in both NSCs and neuroblasts, H3K4me3 is enriched for genes in chromatin modification, neuronal function, and signaling (Supplementary Fig. 6). Our data highlights that the interplay among these biological networks and pathways coordinates adult neurogenesis through H3K4me3 regulation. Intriguingly, the top canonical pathway by IPA prediction from the purified NSCs and combined populations (NSCs and neuroblasts) is "molecular mechanisms of cancer" (p-value: 4.05E-08; Supplementary Fig. 4), in addition to mitochondria function and several signaling pathways, such as eIF4-, RAR-, and Ubiquitin- pathways (Supplementary Fig. 5 and Supplementary Table 1). Although MLL/H3K4me3 overexpression was known to be connected to hematologic diseases[19,20], our finding is the first implication of oncogenic pathways in the SVZ through H3K4me3 regulation.

While H3K4me3 is characteristic of transcriptionally active status in differentiated cells, in embryonic stem cells (ESCs), many genes harboring H3K4me3 are poised for timely activation upon proper differentiation signals[17,18]. Perhaps, the undifferentiated SVZ cells underlying H3K4me3 regulation during adult neurogenesis is reminiscent of similar manner in ESCs. Herein, we seek to determine the functional consequences of H3K4me3 enrichment in terms of expression of its target genes. Through RNA-Seq analysis for undifferentiated SVZ cells purified from baboon brain (Fig. 3A), we found that 72% of H3K4me3 enriched genes in undifferentiated SVZ cells are expressed (1913 genes are detectable by RNA-Seq; 2663 genes enriched with H3K4me3 by ChIP-Seq), in that genes are significantly associated with cell signaling, cellular morphology and maintenance, DNA replication and repair, metabolism, and protein synthesis (Fig. 3H). Our results suggest that H3K4me3 plays potential roles for the integrity of undifferentiated SVZ cells while neurogenesis

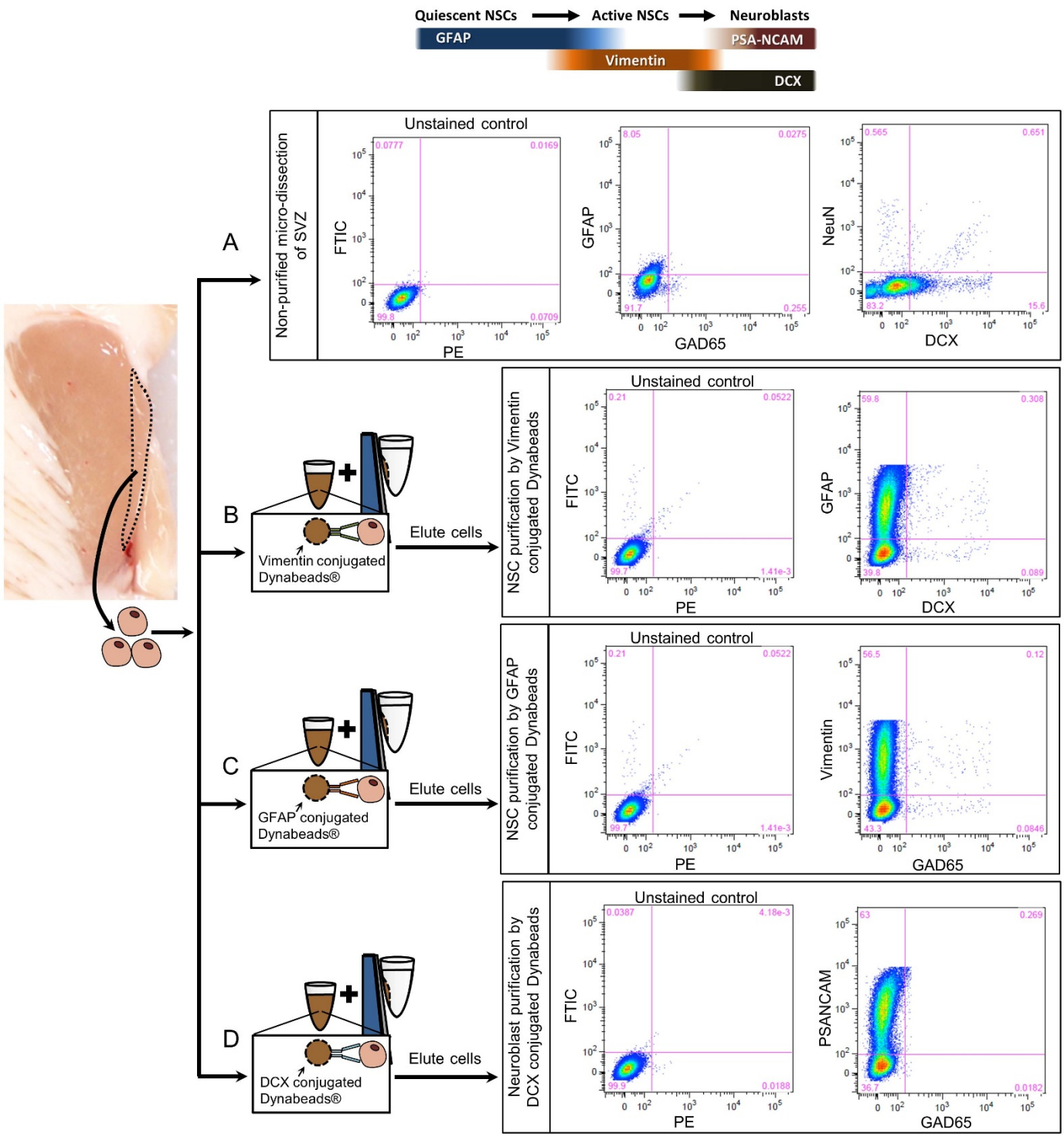

Figure 2 | Flow cytometry analyzes the purity of SVZ cells after dissection and Dynabeads/magnetic isolation. (A) After SVZ dissection, dissociated cells were analyzed by flow cytometry that showed ~8% GFAP-positive (quiescent NSC marker) and 15% DCX-positive (early and migrating neuroblast marker) without contamination of mature cells from nearby caudate (as mature neuron markers NeuN and GAD65 are negative). (B–D) During developing this purification technique, we ran flow cytometry to validate the recovery of NSC and neuroblast cells. (B) For NSCs purification, Vimentin is also a NSC marker, and Dynabeads-Vimentin recovered 59.8% $GFAP^+$ population, but no DCX^+ cells. (C) 56.5% Vimentin-positive cells were recovered from GFAP-Dynabeads purification, supporting the notion that GFAP labels quiescent and active NSCs and Vimentin labels relatively active NSCs. (D) As DCX labels both early and migrating neuroblasts, Dynabeads-DCX recovered 63% PSA-$NCAM^+$ population (PSA-NCAM labels migrating neuroblast), but no $GAD65^+$ cells (GAD65 represents mature neuron marker). Our dissection and purification yield purely undifferentiated SVZ cells without mature cell contamination.

proceeds throughout life time. On the other hand, 750 out of total 2663 H3K4me3-enriched genes are not detectable by RNA-Seq, suggesting 28% of genes in SVZ are poised for activation (Fig. 3H). Lastly, our RNA-Seq identified 11099 genes are expressed in baboon SVZ, whereas only 1913 out of 11099 of genes in undifferentiated SVZ cells are H3K4me3-enriched and also detectable by RNA-Seq

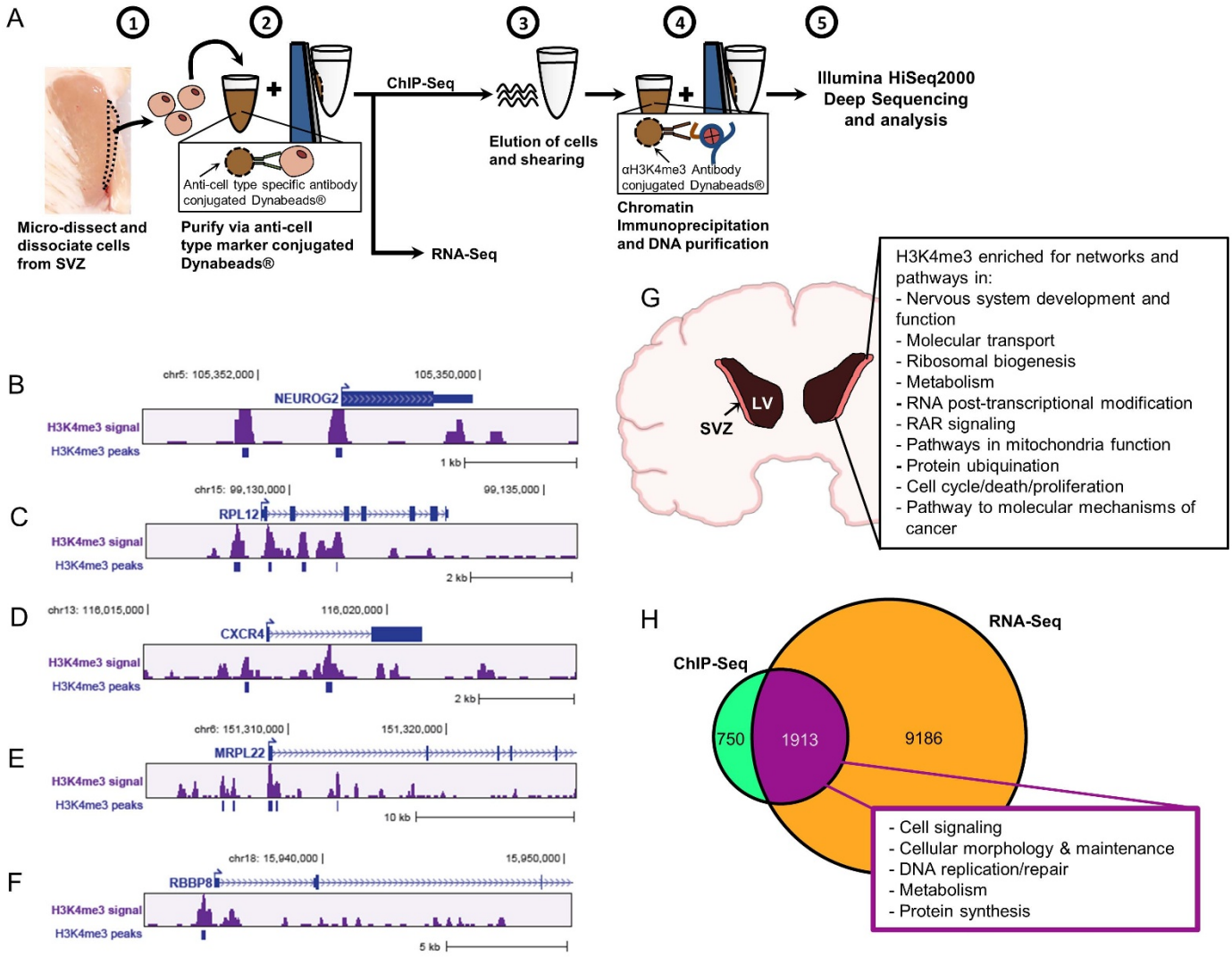

Figure 3 | ChIP-Seq and RNA-Seq analyses for *in vivo* SVZ cells from baboon brain. (A) Scheme displays purification of the *in vivo* SVZ cells for ChIP-Seq and RNA-Seq. For ChIP experiments, DCX- and PSA-NCAM-Dynabeads were used for neuroblast purification, GFAP- and Vimentin-Dynabeads were used for NSC purification. For the pool of undifferentiated SVZ cells, Dynabeads conjugated-GFAP, -Vimentin, -Nestin, -DCX, and -PSANCAM were used for purification. (B–F) Peak views of representative genes enriched with H3K4me3 in baboon SVZ cells. The arrows indicate the annotated transcription start sites (TSS). The H3K4me3 signal is "mapped read density" after normalization to unmodified H3. The H3K4me3 peaks were called at FDR0.01, which is derived from a comparison of mapped read enrichment relative to a local background model based on the binomial distribution. (G) For characterization of H3K4me3 enriched genes in baboon SVZ cells, IPA prediction reveals the top canonical pathways and networks under H3K4me3 regulation in the undifferentiated SVZ cells including NSCs and neuroblasts. (H) RNA-Seq analysis uncovers 11099 genes in the undifferentiated SVZ cells are detectable at transcriptional level, while 1913 expressed genes are enriched with H3K4me3 as noted "magenta" portion.

(Fig. 3H). As the notion of many other factors involved in gene activation, such as opened chromatin through histone acetylation, it is not a surprising result that only ~20% of H3K4me3-enriched genes in undifferentiated SVZ cells are expressed.

Potential link between aberrant H3K4me3 activation and tumorigenesis in subtypes of GBM. The SVZ is a suspected origin for particular types of brain tumor[33–39]. In support of this notion, MRI-classified Glioblastoma Multiforme (GBM) has revealed that group I GBM is arisen from SVZ and group II GBM is associated with SVZ with migratory characteristics[40]. Although in this study, we did not directly address whether group I and II GBM arise from SVZ, our previous findings through proteomic approach identified that markers representing progenitor cells in SVZ including GFAP, Vimentin, and DCX were highly expressed in these GBM tumors[41]. Collectively, i) the close relationships between subtypes of GBM and SVZ; ii) the presence of H3K4me3 in progenitor cells within adult SVZ (Fig. 1); iii) a set of H3K4me3-enriched genes are involved in cancer pathways (Supplementary Fig. 4); and iv) the known association of MLL/H3K4me3 with neoplasia[42,43] prompted us to examine whether the enrichment of H3K4me3 is linked to SVZ-associated GBM. To this end, specimens from MRI-characterized group I and II GBM as well as control regions (Fig. 4A) were analyzed for the relative abundance of H3K4me3 by western blot. The H3K4me3 levels were increased in both group I and II GBM compared to correlated normal brain regions (Fig. 4B). As the pathological consequences of excessive enzymes involved in chromatin modifications were recognized to be harmful, we then examined the expression level of MLL family methyltransferases (MLL1, MLL2, MLL3, MLL4, MLL5) and subunits (Ash2, RbBP5, WDR5) within the core complex responsible for trimethylation of H3K4. There was no difference of MLL3-5 or subunits between the GBM and control specimens, whereas the expressions of MLL1and MLL2 at both transcript and protein levels were increased in group I

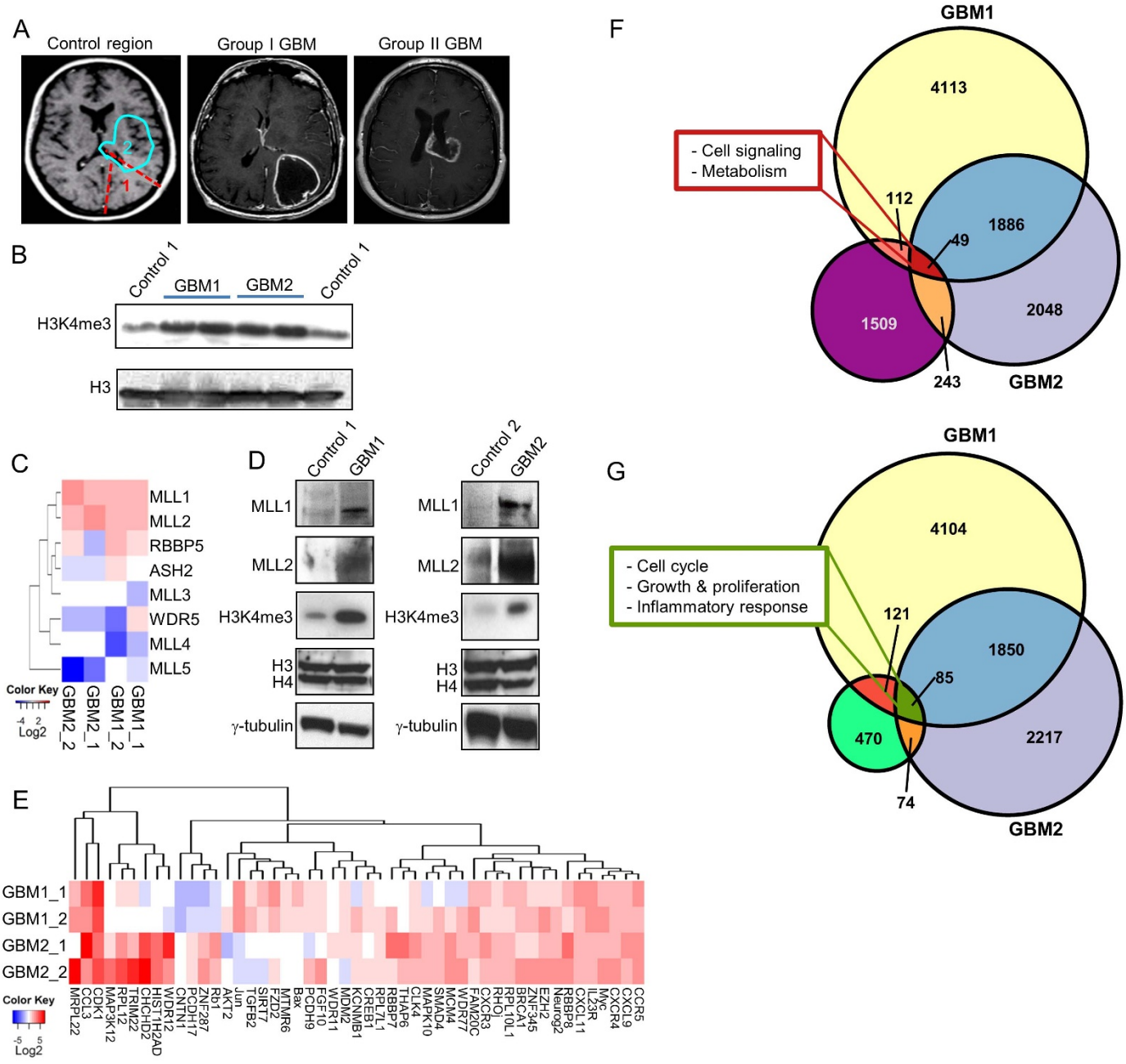

Figure 4 | The levels of H3K4me3, MLL1, MLL2, and subset of H3K4me3-enriched genes are elevated in group I and group II GBM. (A) Scheme displays the control regions and MRI scanned group I and group II GBM. (B) The levels of H3K4me3 were analyzed by western blot. Total histone 3 (H3) were used as loading control. (C) A heatmap for expression of genes in MLL complex (log2 scale) by quantitative RT-PCR for human GBM specimens. (D) Analysis of MLL1 and MLL2 protein levels and H3K4me3 abundance in GBM and control specimens by western blot. γ-tubulin and unmodified H3 and H4 were used as loading controls. (E) A heatmap for expression of selected 49 genes by quantitative RT-PCR for human GBM specimens. Inset shows color scale indicating differences in expression level. The white color with symmetry to 0 defines no difference at expression level between GBM and the controls. The red color scales increased expression level of genes in GBM compared to control and the blue color scales decreased expression in GBM relative to control. The dendrogram was computed by hierarchical clustering with Euclidean distance metric and complete linkage. (F) Venn diagram displays the overlap among genes enriched with H3K4me3 identified from ChIP-Seq of SVZ cells (NSCs and NBs), expressed in SVZ cells identified from RNA-Seq of baboon SVZ cells as noted "magenta" portion (n = 1913), and substantially increased in SVZ-associated GBM cases identified from RNA-Seq analyses (yellow- group I GBM; light purple- group II GBM). The GO terms for overrepresented genes among this overlap are shown in the "box". (G) Venn diagram displays the overlap among genes which are not detectable in baboon SVZ cells by RNA-Seq as noted "green" portion (n = 750), but are enriched with H3K4me3 identified from ChIP-Seq of baboon SVZ cells (NSCs and NBs), and are substantially increased in SVZ-associated GBM cases by RNA-Seq analyses (yellow- group I GBM; light purple- group II GBM). The GO terms for overrepresented genes among this overlap are shown in the "box".

and II GBM compared to control (Fig. 4C, D). Herein, we envision the elevated level of H3K4me3 in GBM could be due to the up-regulation of MLL1 and MLL2. In light of the "molecular mechanisms of cancer" as the top canonical pathway from our H3K4me3 ChIP-Seq data raises the possibility that aberrant H3K4me3 landscape may perturb SVZ lineages and contribute to

tumorigenesis. To examine whether genes enriched by H3K4me3 in the primate SVZ cells were consequently affected in the group I and II GBM specimens, we applied RNA-Seq for specimens obtained from these types of GBM and control brain region (Fig. 4A) to delineate the differential gene expression in group I and II GBM at the genome-wide scale. Of note, the overlap between these 2 types of GBM includes genes functioning in chromatin remodeling, cell cycle, metabolism, signaling, and ribogenesis. We validated RNA-Seq data by qRT-PCR for selective genes from these categories. After normalization to control brain regions, several genes involved in chromatin modification, ribosomal biogenesis, and metabolism were increased in both group I and II GBM (Fig. 4E), resembling our previous proteomic results[41]. Consistent with RNA-Seq results, a number of genes were up-regulated specifically in either group I or II GBM by qRT-PCR validation. These differential expression signatures may be useful for distinguishing the molecular characteristics between group I and II GBM. H3K4me3 is substantially increased in group I and II GBM (Fig. 4B and D), implying that these tumors may have a transcriptional profile reflecting the H3K4me3 enrichment in progenitor cells of the SVZ. To further examine this relationship, we compared the overlap of RNA-Seq data between SVZ-associated GBM and undifferentiated SVZ cells from baboon brain (Supplementary Table 2). We found that 404 genes detectable by RNA-Seq in SVZ and enriched with H3K4me3 were aberrantly overexpressed ($\log2 \geq 2$) in group I and/or group II GBM, respectively (Fig. 4F and Supplementary Table 2). These genes are significantly associated with cell signaling, cellular morphology and maintenance, DNA replication and repair. Additionally, 49 genes involved in cell signaling and metabolism are expressed and enriched with H3K4me3 in undifferentiated SVZ cells, and are also upregulated in both group I and group II GBM (Fig. 4F). Intriguingly, a set of H3K4me3-enriched genes in undifferentiated baboon SVZ cells are not detectable by RNA-Seq, but are overexpressed in GBM ($n = 85$) (Fig. 4G and Supplementary Table 2). These 85 genes predominantly function in cell cycle, cellular growth and proliferation, and immune response, as such reflects to the profound effect of aberrant H3K4me3 activation that causes proliferating cancer cells. Perhaps, these mitotic and proliferation signatures would be a better tracing of cancer cells within the surrounding SVZ cells.

Discussion

To glean previously unattainable information regarding the epigenetic regulation of H3K4me3 in the *in vivo* progenitor cells within the primate SVZ, we undertook the genome-wide ChIP-Seq and RNA-Seq approaches for purified SVZ cells from baboon brain. Using ChIP-Seq, a cohort of H3K4me3-enriched genes was identified, which function in chromatin modification, cell cycle, cell signaling, metabolism, nervous system differentiation, ribosomal biogenesis, and protein synthesis. As H3K4me3 has been linked to poised or active gene expression[15,17,18], our RNA-Seq analysis reveals the detectable expression from a subset of H3K4me3-enriched genes that contribute to cellular signaling, DNA replication, metabolism, and protein synthesis. Our findings illustrate the importance of H3K4me3 in coordinating cell fates, metabolic sensing, and signaling pathways for life-long neurogenesis. It is worth noticing in the independent set of ChIP-Seq for the purified NSCs that H3K4me3 is preferentially enriched for genes involved in cell cycle and DNA replication/repair, implicating that genomic integrity in NSCs is maintained under H3K4me3 regulation. Further, we showed that H3K4me3 is predominantly enriched for genes responsible for ribogenesis in the purified neuroblast population. Remarkably, a substantial number of H3K4me3-enriched genes in both NSCs and neuroblasts are linked to cancer pathways.

While MLL/H3K4me3 functions in adult neurogenesis, the suspicion of SVZ as the origin for subtypes of brain tumors[38,40] raises the concern that, similar to hematologic cancer[19,20], deregulation of this epigenetic mark could cause cancer associated with this germinal niche. We found elevated levels of H3K4me3 in MRI-classified group I and II GBM specimens, reminiscent of a recent study of chromatin status mapping, in that enriched occupancy of H3K4me3 was found in established GBM cell lines[44]. Although previous studies have identified mutations and the resulting abnormal expression (eg. p53, IDH1, IDH2) associated with distinct types of GBM[45–47], our RNA-Seq result showed that the expression of those genes were not affected, suggesting the SVZ-associated group I and II GBM is distinct from other characterized GBM cases. Further integrated analyses across ChIP-Seq (baboon SVZ cells) and RNA-Seq (baboon SVZ cells; group I GBM; group II GBM) data revealed that the expression level of 16.7% of H3K4me3-enriched genes is elevated in group I and II GBM, regardless whether the transcripts are detectable or undetectable in the *in vivo* baboon SVZ cells. Among those genes, a significant overrepresentation of GO terms is related to cell cycle, proliferation, and signaling for immune responses, which imply the most profound effect of H3K4me3 in these SVZ-associated GBM. Our findings also suggest that other factors existing in tumor microenvironment could initiate or enhance tumorigenesis in addition to aberrant accumulation of H3K4me3. However, changes in subsets of gene expression under aberrant H3K4me3 activation may, at least in part, influence cell fate in the SVZ and disrupt the maintenance of neurogenesis that ultimately contribute to tumorogenesis in subtypes of GBM. Apparently, there is a fine line that H3K4me3 must walk between adult neurogenesis and oncogenesis in the SVZ. Thus, future studies to understand the mechanism underlying the fine tuning regulation of H3K4me3 in SVZ cells remains to be determined.

Methods

All animal experiments were approved by the guidelines of the Institutional Animal Care and Use Committee of the University of Texas at San Antonio (UTSA) and Texas Biomedical research Institute/Southwest National Primate Research Center (SNPRC) at San Antonio. The experimental materials involving human specimens were approved by the Institutional Review Board (IRB) before starting research.

Cell types purification for ChIP-Seq analysis. We manually conjugated antibody against cell type markers, such as GFAP, Nestin, Vimentin, PSA-NCAM, or Doublecortin to Dynabeads. We then used the Dynabeads-conjugated antibody to purify cells dissociated from SVZ microdissection. Briefly, cells from fresh dissected baboon SVZ were immediately dissociated with Accutase, equilibrated in binding buffer containing phosphate-buffered saline (PBS), 0.05% TritonX-100 (or saponin, detergent choice depends upon antibody), and subsequently subjected to Dynabeads-conjugated antibody purification. After elution with high salt and pH-gradient buffer, the purified populations were crosslinked in 1.1% formaldehyde before chromatin shearing by Diagenode Bioruptor. The resulting sheared chromatin fragments in a size range between 200 to 500 base pairs were incubated with H3K4me3 antibody-conjugated Protein A Dynabeads (Millipore #07-473,1 : 1000; Active Motif #39160, 1 : 500; Life Technology Dynabeads protein A) overnight. For normalization, the aliquot of sheared chromatin fragments were incubated with antibody against total histone 3-conjugated Protein A Dynabeads (unmodified H3 antibody, Millipore #05-499; 1 : 1000). Subsequently, enriched chromatin fragments were eluted, subjected to de-crosslink, purified for library preparation (Illumina Library Kit), and deep-sequencing with 200 million tags through Illumina HiSeq2000 sequencer.

Sequence alignment and peak calling. The Rhesus macaque gene annotation was constructed and is maintained at the UCSC and is derived directly from the NCBI RefSeq project (http://nar.oxfordjournals.org/content/33/suppl_1/D501.full). Alignments were generated by using the Bowtie alignment program version 0.12.7., with a maximum of 2 mismatches allowed in the retained uniquely mapping reads. Aligned read enrichments where detected using the Hotspot peak finding program that is based upon a binomial distribution model, with built-in, automatic background adjustment. Hotspot to RefSeq gene associations were established in a distance ranked manner by using the closest-features program from the BEDOPS tool set[32] (https://code.google.com/p/bedops/).

RNA-Seq analysis. Total RNA was extracted from purified baboon SVZ cells and frozen GBM specimens using TRIzol reagent and sequencing libraries were generated with Illumina RNA-Seq library kit. Paired-end RNA-deepSeq (76 base pair; >300 million tag reads) were aligned to hg19. DESeq was used to normalize raw read counts; and Cufflink reports read counts and estimated FPKM (fragments per kilobase of exon per million fragments mapped; http://cufflinks.cbcb.umd.edu/faq.html#fpkm). Genes with expression values >1 FPKM were considered for subsequent

analyses. Cuffdiff was applied to analyze the differential gene expression between GBM and control. Area-proportional Venn Diagrams were generated using the R package "Vennerable" (http://r-forge.r-project.org/projects/vennerable/).

GO, network, and pathway analysis. Gene ontology analysis was performed using the DAVID functional analysis tool[48]. The Bonferroni, Benjamini, and FDR (false discovery rate) were used for multiple test correction. Functional pathway and network analysis of bound loci were performed using Ingenuity Pathway Analysis (IPA) (Ingenuity® Systems, Redwood City, CA, USA). Loci, known as focus genes, were grouped into ontology classes. The Ingenuity Knowledge Base, a repository of biological and chemical interactions, was used as a reference set for pathway and network inference. Over-represented signaling and metabolic canonical pathways in the input data were determined based on two parameters: 1) The ratio of the number of molecules from the focus loci set that map to the pathway divided by the total number of molecules that map to the canonical pathway, and 2) a *p* value calculated by Fisher's exact test that determines the probability that the association between the focus loci and the canonical pathway is explained by chance alone. Network analysis used focus loci as "seeds" to infer *de novo* interaction networks. Direct or indirect interactions between focus loci and other molecules were inferred based on experimentally observed relationships supported by at least one reference from the literature. Additional molecules from the Ingenuity Knowledge Base were added to the network to fill or join smaller networks. The network score was based on the hypergeometric distribution and calculated with the right-tailed Fisher's exact test. A higher score indicates a lower probability of finding the observed number of focus molecules in a given network by chance.

Co-immunostaining and confocal imaging. *EdU tracing*: Mice were administered with EdU (150 μg/body weight) by IP, and subsequently perfused/fixed with 1X PBS and 4% Paraformaldehyde 2 hours after EdU injection. Brain sections were processed and permeabalized with 0.5% Triton X-100 for 20 minutes at room temperature prior to the detection assay of EdU incorporation under manufacture instruction (Invitrogen Click-it Edu imaging kit: copper-catalyzed covalent reaction between the Edu alkyne and the Alexa Fluor© dye azide). Edu was labeled by Alexa Fluor© 488 and imaged under Zeiss LSM 510 Fluorescent Confocal microscope. AraC antimitotic treatment was performed as described previously[41]. Mouse brains were OCT embedded and then 12 μm frozen sections were processed for immunostaining with antibodies against Doublecortin (Cell Signaling #4604; 1 : 500), Polysialic Acid-NCAM, clone 2-2b (Millipore #MAB5324, Lot# 1966892; 1 : 500), Glial Fibrillary Acidic Protein, clone GA5 (Millipore #MAB3402, Lot#1993774; 1 : 500), MASH1(Abcam #ab38556; 1 : 500), H3K4me3 (Millipore #07-473; 1 : 500), EdU 5-ethynyl-2′-deoxyuridine (Invitrogen #C10337), and Vimentin (Sigma V2258;1 : 500). Coronal slices of fresh baboon forebrain were fixed in 4% paraformaldehyde overnight, cryo-protected with 30% sucrose/PBS prior to OCT embedding and sectioning (60 μm). Antibodies used for co-immunostaining are described above in addition to Sox2 (Cell Signaling #3579 and Millipore #AB5603; 1 : 500) and Nestin (Abcam #AB27952; 1 : 500). Fluorescent labeling with secondary antibodies AlexaFIuor 488 (Molecular Probes, dilution 1 : 1000) and AlexaFluor 594 (Molecular Probes 1 : 1000) were acquired by confocal microscope. Vectashield with DAPI (Vector Laboratories Ltd, # H-1200) was used for mounting medium and counter stain. Baboon SVZ imaging at 40× focal stacks and 40×/100× 2um single slice were taken on a Zeiss 710 two-photon confocal microscope under the control of the Zeiss Zen 2011 software (Carl Zeiss Microscopy; Germany). Mouse SVZ imaging at 40× 2um single slice was taken on a Zeiss 510 confocal microscope.

Western blot analysis of human brain tissue samples. The snap-frozen brain tissue samples were homogenized in 1 ml Buffer A (10 mM Tris-HCl pH = 8, 250 mM Sucrose, 10 mM MgCl2, 1 mM EGTA) with 1X protease inhibitor to extract cytoplasmic proteins. The resulting pellets were further homogenized in 1 ml RIPA buffer containing 1X protease inhibitor to isolate nuclear fraction. The total protein concentrations in cytoplasmic and nuclear fractions were quantified by Bradford assay (Bio-Rad). Bulk chromatin extraction was performed with histone extraction kit from Active Motif. For western blot, equal amounts of protein from normal or GBM specimens were denatured in final 1X SDS stop buffer and subjected to SDS-PAGE for western blot analysis with antibodies against H3K4me3 (Millipore #07-473; 1 : 1000), H3, H4, MLL1, MLL2, MLL3, and γ-tubulin (Sigma #T5326; 1 : 1000). Subsequently, HRP-conjugated secondary IgG (Cell Signalling; 1 : 2000) and enhanced chemiluminescence kit (ECL plus; GE) were used for detection. Of note, blotting membranes were cut into strips to accommodate the molecular weights of MLL, H3K4me3, and γ-tubulin for antibody incubation separately. Anti-H3K4me3 was stripped after ECL detection and the blots were subsequently incubated with unmodified H3 and H4 antibodies (internal controls).

Gene expression analysis. Total RNA was extracted using TRIzol reagent (Invitrogen) and cDNA was synthesized by cDNA synthesis kit (Applied Biosystems) according to manufacturer's instruction. The expression level was measured by Real-Time PCR (RT-PCR) with Power SYBR Green detection on ABI7900HT system (SDS 2.4, Applied Biosystems). S16 was used as an internal control for normalization (ΔCt). The expression level was measured by ΔΔCt (log2 scale) after subtraction of ΔCt between GBM and control. A heatmap was generated using the heatmap2 function in R. A hierarchical clustering algorithm with Euclidean distance metric and complete linkage was implemented to generate the hierarchical tree.

Flow cytometry. Isolated SVZ cells were incubated on ice for 10 minutes with anti-mouse CD16/CD32 (clone 2.4G2; BD Pharmingen) to block FcRs. Cells were then incubated on ice for 30 min with primary antibodies (anti-GFAP, Vimentin, GAD65, NeuN, PSA-NCAM, and Doublecortin), and subsequently labeled with Alexa fluor 488- and PE-conjugated secondary antibodies (BD Biosciences). Primary antibody resources: Doublecortin (Cell Signaling #4604; Millipore clone2G5 MABN707), Polysialic Acid-NCAM, clone 2-2b (Millipore #MAB5324, Lot# 1966892), Glial Fibrillary Acidic Protein, clone GA5 (Millipore #MAB3402, Lot#1993774; Pierce #MA1-35376; Abcam clone 2A5 #ab4648), Sox2 (Cell Signaling #3579 and Millipore #AB5603; 1 : 500), Nestin (Abcam #AB27952; 1 : 500), NeuN (Millipore MAB377), and GAD65 (Abcam #AB11070). For analysis, controls including the positive controls stained with each antibody separately, the isotype controls, and the unstained cells were used for gate compensation. Flow data was acquired on a LSR-II (BD Biosciences) configured with an argon 488 laser with a 505 LP dichroic and 525/50 filter to detect Alexa fluor 488 and a green 510 laser with a 735 LP dichroic and a 575/26 filter to detect PE. Compensation and data analysis was performed using FlowJo software (Tree Star, Inc, Ashland, OR).

1. Ming, G. L. & Song, H. Adult neurogenesis in the mammalian brain: significant answers and significant questions. *Neuron* **70**, 687–702 (2011).
2. Ihrie, R. A. & Alvarez-Buylla, A. Cells in the astroglial lineage are neural stem cells. *Cell Tissue Res* **331**, 179–91 (2008).
3. Clelland, C. D. *et al.* A functional role for adult hippocampal neurogenesis in spatial pattern separation. *Science* **325**, 210–3 (2009).
4. Jessberger, S. *et al.* Epigenetic modulation of seizure-induced neurogenesis and cognitive decline. *J Neurosci* **27**, 5967–75 (2007).
5. Alvarez-Buylla, A. & Lim, D. A. For the long run: maintaining germinal niches in the adult brain. *Neuron* **41**, 683–6 (2004).
6. Ihrie, R. A. & Alvarez-Buylla, A. Lake-front property: a unique germinal niche by the lateral ventricles of the adult brain. *Neuron* **70**, 674–86 (2011).
7. Fuentealba, L. C., Obernier, K. & Alvarez-Buylla, A. Adult neural stem cells bridge their niche. *Cell Stem Cell* **10**, 698–708 (2012).
8. Doetsch, F., Petreanu, L., Caille, I., Garcia-Verdugo, J. M. & Alvarez-Buylla, A. EGF converts transit-amplifying neurogenic precursors in the adult brain into multipotent stem cells. *Neuron* **36**, 1021–34 (2002).
9. Zheng, W., Nowakowski, R. S. & Vaccarino, F. M. Fibroblast growth factor 2 is required for maintaining the neural stem cell pool in the mouse brain subventricular zone. *Dev Neurosci* **26**, 181–96 (2004).
10. Jackson, E. L. *et al.* PDGFR alpha-positive B cells are neural stem cells in the adult SVZ that form glioma-like growths in response to increased PDGF signaling. *Neuron* **51**, 187–99 (2006).
11. Ihrie, R. A. *et al.* Persistent sonic hedgehog signaling in adult brain determines neural stem cell positional identity. *Neuron* **71**, 250–62 (2011).
12. Molofsky, A. V. *et al.* Increasing p16INK4a expression decreases forebrain progenitors and neurogenesis during ageing. *Nature* **443**, 448–52 (2006).
13. Kohwi, M., Osumi, N., Rubenstein, J. L. & Alvarez-Buylla, A. Pax6 is required for making specific subpopulations of granule and periglomerular neurons in the olfactory bulb. *J Neurosci* **25**, 6997–7003 (2005).
14. Lim, D. A. *et al.* Chromatin remodelling factor Mll1 is essential for neurogenesis from postnatal neural stem cells. *Nature* **458**, 529–33 (2009).
15. Wysocka, J. *et al.* WDR5 associates with histone H3 methylated at K4 and is essential for H3 K4 methylation and vertebrate development. *Cell* **121**, 859–72 (2005).
16. Ruthenburg, A. J., Allis, C. D. & Wysocka, J. Methylation of lysine 4 on histone H3: intricacy of writing and reading a single epigenetic mark. *Mol Cell* **25**, 15–30 (2007).
17. Bernstein, B. E. *et al.* A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**, 315–26 (2006).
18. Voigt, P., Tee, W. W. & Reinberg, D. A double take on bivalent promoters. *Genes Dev* **27**, 1318–38 (2013).
19. Yokoyama, A. *et al.* The menin tumor suppressor protein is an essential oncogenic cofactor for MLL-associated leukemogenesis. *Cell* **123**, 207–18 (2005).
20. Shen, H. & Laird, P. W. Interplay between the Cancer Genome and Epigenome. *Cell* **153**, 38–55 (2013).
21. Baylin, S. B. & Jones, P. A. A decade of exploring the cancer epigenome - biological and translational implications. *Nat Rev Cancer* **11**, 726–34 (2011).
22. Pujadas, E. & Feinberg, A. P. Regulated noise in the epigenetic landscape of development and disease. *Cell* **148**, 1123–31 (2012).
23. Sanai, N. *et al.* Unique astrocyte ribbon in adult human brain contains neural stem cells but lacks chain migration. *Nature* **427**, 740–4 (2004).
24. Sanai, N. *et al.* Corridors of migrating neurons in the human brain and their decline during infancy. *Nature* **478**, 382–6 (2011).
25. Arellano, J. I. & Rakic, P. Neuroscience: Gone with the wean. *Nature* **478**, 333–4 (2011).
26. Kochunov, P. *et al.* Genetics of primary cerebral gyrification: Heritability of length, depth and area of primary sulci in an extended pedigree of Papio baboons. *Neuroimage* **53**, 1126–34 (2010).
27. Rogers, J. *et al.* On the genetic architecture of cortical folding and brain volume in primates. *Neuroimage* **53**, 1103–8 (2010).

28. Doetsch, F., Caille, I., Lim, D. A., Garcia-Verdugo, J. M. & Alvarez-Buylla, A. Subventricular zone astrocytes are neural stem cells in the adult mammalian brain. *Cell* **97**, 703–16 (1999).
29. Doetsch, F., Garcia-Verdugo, J. M. & Alvarez-Buylla, A. Regeneration of a germinal layer in the adult mammalian brain. *Proc Natl Acad Sci U S A* **96**, 11619–24 (1999).
30. Gibbs, R. A. *et al.* Evolutionary and biomedical insights from the rhesus macaque genome. *Science* **316**, 222–34 (2007).
31. John, S. *et al.* Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. *Nat Genet* **43**, 264–8 (2011).
32. Neph, S. *et al.* BEDOPS: high-performance genomic feature operations. *Bioinformatics* **28**, 1919–20 (2012).
33. Ligon, K. L. *et al.* Olig2-regulated lineage-restricted pathway controls replication competence in neural stem cells and malignant glioma. *Neuron* **53**, 503–17 (2007).
34. Po, A. *et al.* Hedgehog controls neural stem cells through p53-independent regulation of Nanog. *EMBO J* **29**, 2646–58 (2010).
35. Stiles, C. D. & Rowitch, D. H. Glioma stem cells: a midterm exam. *Neuron* **58**, 832–46 (2008).
36. Zbinden, M. *et al.* NANOG regulates glioma stem cells and is essential in vivo acting in a cross-functional network with GLI1 and p53. *EMBO J* **29**, 2659–74 (2010).
37. Zheng, H. *et al.* PLAGL2 regulates Wnt signaling to impede differentiation in neural stem cells and gliomas. *Cancer Cell* **17**, 497–509 (2010).
38. Chen, J., McKay, R. M. & Parada, L. F. Malignant glioma: lessons from genomics, mouse models, and stem cells. *Cell* **149**, 36–47 (2012).
39. Chen, J. *et al.* A restricted cell population propagates glioblastoma growth after chemotherapy. *Nature* **488**, 522–6 (2012).
40. Lim, D. A. *et al.* Relationship of glioblastoma multiforme to neural stem cell regions predicts invasive and multifocal tumor phenotype. *Neuro Oncol* **9**, 424–9 (2007).
41. Haskins, W. E. *et al.* Molecular Characteristics in MRI-Classified Group 1 Glioblastoma Multiforme. *Front Oncol* **3**, 182 (2013).
42. Feinberg, A. P., Ohlsson, R. & Henikoff, S. The epigenetic progenitor origin of human cancer. *Nat Rev Genet* **7**, 21–33 (2006).
43. Luscher, B. & Larsson, L. G. The basic region/helix-loop-helix/leucine zipper domain of Myc proto-oncoproteins: function and regulation. *Oncogene* **18**, 2955–66 (1999).
44. Rheinbay, E. *et al.* An aberrant transcription factor network essential for wnt signaling and stem cell maintenance in glioblastoma. *Cell Rep* **3**, 1567–79 (2013).
45. Verhaak, R. G. *et al.* Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell* **17**, 98–110 (2010).
46. Phillips, H. S. *et al.* Molecular subclasses of high-grade glioma predict prognosis, delineate a pattern of disease progression, and resemble stages in neurogenesis. *Cancer Cell* **9**, 157–73 (2006).
47. Liu, C. *et al.* Mosaic analysis with double markers reveals tumor cell of origin in glioma. *Cell* **146**, 209–21 (2011).
48. Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44–57 (2009).

Acknowledgments

We acknowledge the UCSF, Department of Neurological Surgery, Brain Tumor Research Center tissue core as the resource of the human specimens. Sequencing was performed at the High Throughput Genomics Center in the Department of Genome Sciences, University of Washington, Seattle, WA. The authors are grateful to Dr. Gary Gaufo at UTSA for usage of the LSM 710; Daniel Bates, Jenny Garcia, Kaifeng Hung, Gabriel Leal, and Bethany Zablotsky for their technical support; as well as Drs. Joseph Costello and Arturo Alvarez-Buylla at UCSF for critical discussion. We thank Texas Biomedical Research Institute/Southwest National Primate Research Center (SNPRC) at San Antonio for baboon tissues and Dr. Jera Pecotte for technical support. The non-human primate project used biological materials were funded by the National Center for Research Resources (p51 RR013986) and the Office of Research Infrastructure Programs/OD P51 OD011133. This work received support from Computational System Biology Core and Imaging Core, funded by the National Institute on Minority Health and Health Disparities (G12MD007591) from the National Institutes of Health. DAG and MRF were supported by scholarship from UTSA, Department of Biology PhD graduate program. YW is supported by NIH grants GM100806 and AI080579. This project is supported by the SPORE Grant 5 P50 CA097257-10 to MSB, as well as the 5P30GM092334 from the National Institutes of Health, the UTSA Institutional Start-up fund, and TRAC award (Tenure-track Research Award Competition) to CHL.

Author contributions

R.S.S. performed deep sequencing and alignment/peak calls analyses; M.R.F. performed most of imaging for Fig. 1, performed protein extraction and western blot, made all schematics and isolation schemes; D.A.G. performed part of imaging for Fig. 1; E.H. performed RNA-Seq analysis; C.T.R. performed overlap comparison of genome-wide data and assisted flow cytometry; C.F.P. performed baboon brain necropsy; A.E.C. provided flow cytometry analysis; Y.W. performed GO/network/pathway and gene expression data analyses; M.S.B. provided GBM samples and MRI analyses; C.H.L. designed study and experiments, performed ChIP, and wrote manuscript; R.S.S. and M.S.B. provided advice throughout the work; All authors reviewed manuscript.

Additional information

Supplementary information accompanies this paper at http://www.nature.com/scientificreports

Accession code: All ChIP-Sequencing and RNA-Sequencing data generated during the course of this study have been deposited to GEO under the accession ID GSE#58531.

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Sandstrom, R.S. *et al.* Epigenetic Regulation by Chromatin Activation Mark H3K4me3 in Primate Progenitor Cells within Adult Neurogenic Niche. *Sci. Rep.* **4**, 5371; DOI:10.1038/srep05371 (2014).