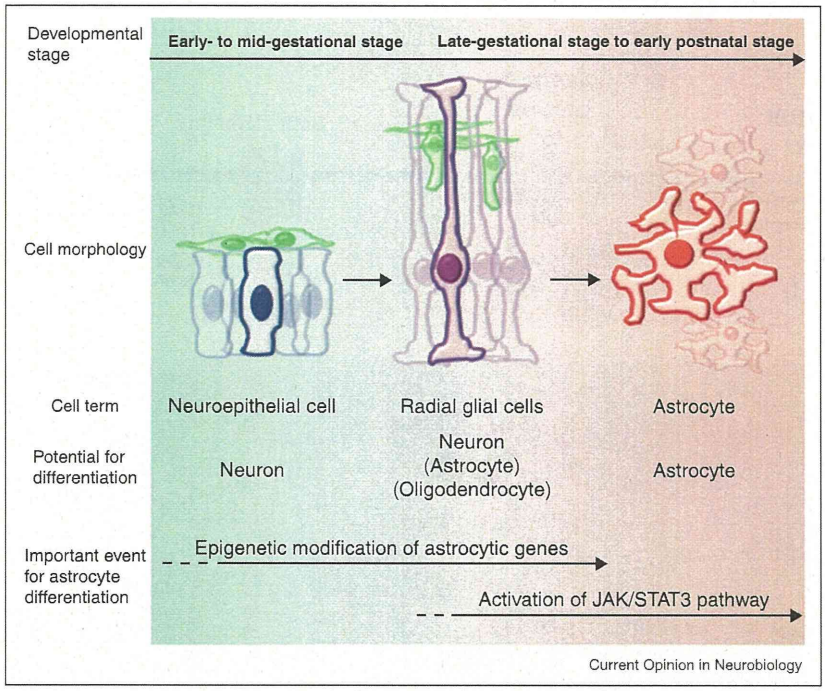
Figure 1



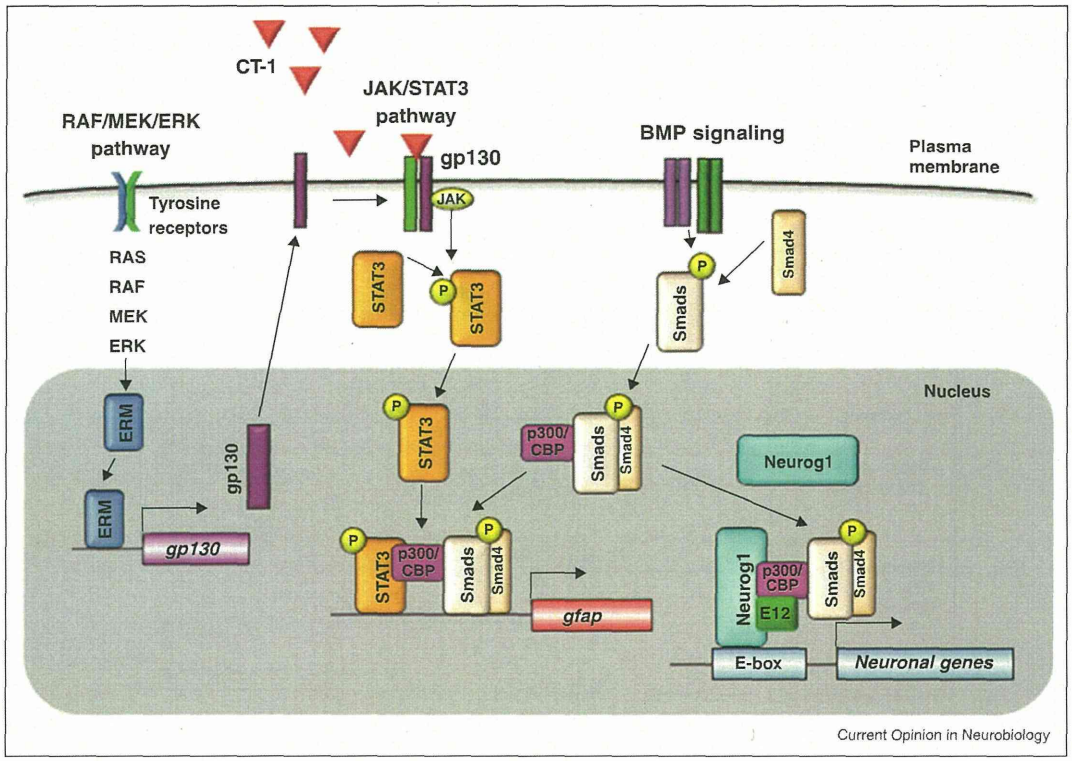
Temporal development of neural stem cells (NSCs), exemplified by the ventricular zone of the cerebral cortex. In early-gestational to mid-gestational stages (left), NSCs, referred to as neuroepithelial cells (blue), divide symmetrically and asymmetrically to increase their own numbers and to generate early neurons (green). As development proceeds (middle), the morphology of NSCs changes to have and the cells acquire radial processes extending to the pial surface from their somata, which reside in proximity to the ventricular boundaries; the cells are now referred to as radial glial cells (violet). These NSCs also generate many neurons (green) but have the additional potential to differentiate into astrocytes and oligodendrocytes. At the late-gestational and early postnatal stages (right), most NSCs begin to detach from the apical side and migrate into the cortex, and differentiate into astrocytes (red). In this scheme, there are two important events for astrocytogenesis. The first is epigenetic derepression of astrocytic gene transcription for the acquisition of astrocyte differentiation potential by NSCs, which is sustained until late embryonic stages; the second is activation of cytokine signaling through the JAK/STAT3 pathway for the induction of astrocytic gene transcription.

transcription factors form a complex with Smad4 which is translocated into the nucleus to activate the transcription of specific genes. Smads are known to participate in the induction of astrocytic gene expression by forming a complex with STAT3 that is bridged by the transcriptional coactivators p300/CBP [10].

Deletions of the genes encoding LIF [11], LIFRβ [12], gp130 [13] or STAT3 [14] all result in impaired astrocyte differentiation *in vivo*, indicating that the JAK–STAT3 pathway is essential for astrocytogenesis in the developing brain. It has also been reported that CT-1 secreted from neurons is important for astrocyte differentiation from NSCs in embryonic stages [15]. Furthermore, the proneuronal transcription factor Neurogenin 1 (Neurog1) may suppress astrocyte differentiation from NSCs by sequestering the p300/CBP-Smads complex away from STAT3, leading to the suppression of STAT3 target genes [16,17] (Figure 2). Neurog1 is expressed during the neurogenic period, but not the astrocytogenic period, of neocortical development. Its expression is precisely controlled by polycomb group proteins through chromatin modification [18], thus contributing to the mechanisms for fate-switching of NSCs from neurogenic to astrocytogenic as development proceeds.

Recently, the RAF/MEK/ERK pathway, which is among the best-characterized signaling cascades in mediating the effects of extracellular factors, has been shown to participate in astrocytogenesis from NSCs by modulating the JAK–STAT3 pathway. Li et al. [19•] reported that mutation of the core RAF/MEK/ERK pathway components MEK1 (MAP2K1) and MEK2 (MAP2K2) in NSCs caused a severe reduction in the number of astrocyte precursor cells expressing the aldehyde dehydrogenase 1 family member L1 [20] in late-gestational forebrain. They also found that the JAK/STAT3 pathway in NSCs derived from *Mek1/2* null mutant embryos was attenuated due to a marked reduction in the expression of gp130, which is critical for activation of the pathway by IL-6

Figure 2



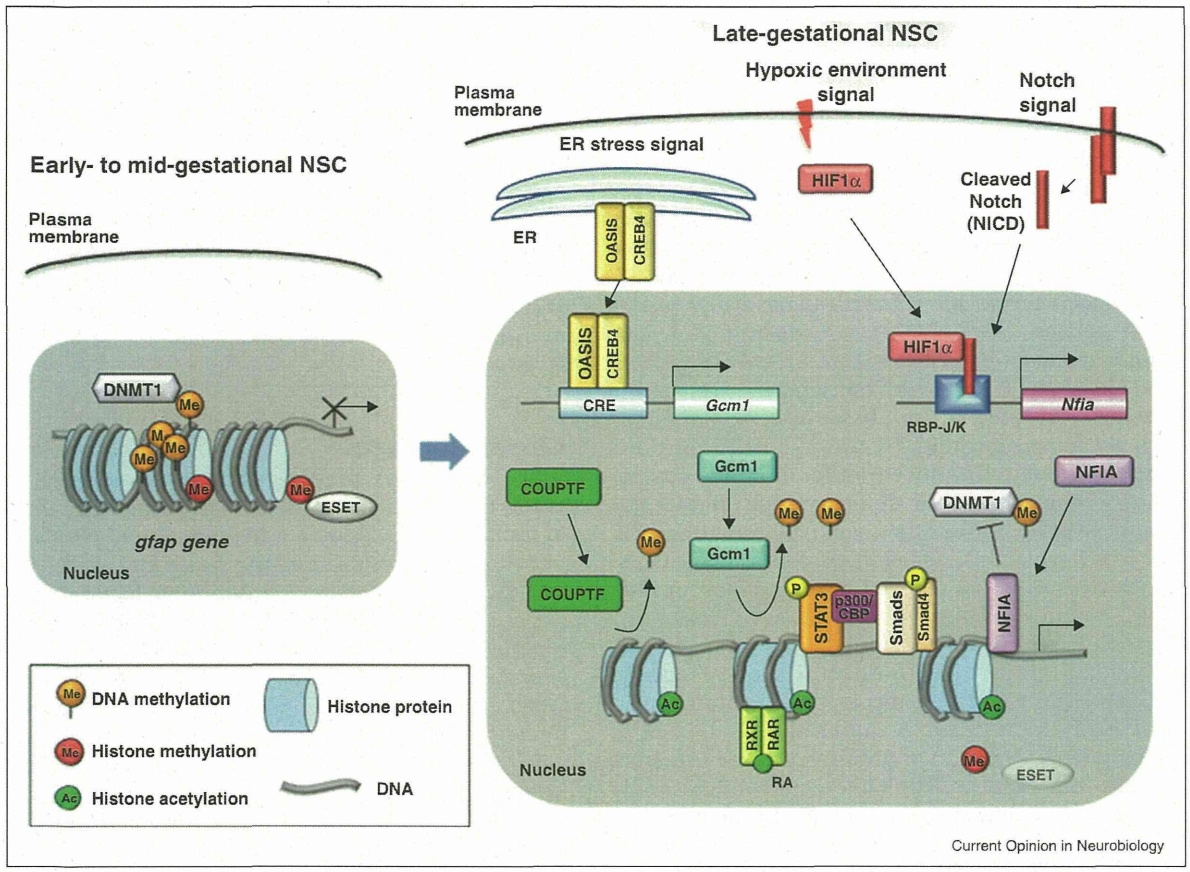
Regulators modulating the signaling pathway of astrocyte-inducing cytokines. Synergistic integration of the JAK/STAT3 pathway and BMP/Smads signaling is achieved by the formation of a complex involving their respective downstream transcription factors, STAT3 and Smads, together with p300/CBP. Neuron-secreted CT-1 is important for activation of the JAK/STAT3 pathway in NSCs during embryonic stages. Neurog1 sequesters the p300/CBP-Smads complex away from STAT3, leading to the suppression of STAT3 target genes. The RAF/MEK/ERK pathway also plays an important role in astrocytogenesis from NSCs by regulating gp130 expression.

family cytokines. In addition, Li et al. identified the Ets transcription factor family member Etv5/Erm as a downstream target of MEKs in NSCs [19•]. Overexpression of Etv5/Erm could rescue the impairment of cytokine-induced differentiation into astrocytes in MEK null mutant NSCs, suggesting that Etv5/Erm regulates gp130 expression in NSCs. Collectively, these observations indicate that the RAF/MEK/ERK pathway plays an important role in astrocytogenesis from NSCs through the regulation gp130 expression (Figure 2).

Epigenetic modifiers of astrocytic genes in NSCs

Although it is clear that activation of the JAK–STAT3 pathway by IL-6 family cytokines induces astrocyte differentiation from NSCs, as described above, early-gestational and mid-gestational-stage NSCs are insensitive to these cytokines in terms of astrocyte differentiation; in other words, astrocyte differentiation is inhibited in NSCs during these developmental stages [21]. This inhibition is known to be achieved by a particular epigenetic modification, DNA methylation. Since the promoters of astrocytic genes such as *gfap* are highly methylated in NSCs at early-gestational and mid-gestational stages, STAT3 binding to these promoters is impeded (Figure 3). As gestation proceeds, astrocytic gene promoters are demethylated and the NSCs then become competent to express these genes in response to astrocyte-inducing cytokines, suggesting that DNA methylation plays an important role in defining the timing of astrocytogenesis from NSCs during development [21,22]. We have demonstrated previously that Notch signaling and its downstream target nuclear factor IA (NFIA) play a crucial role in the demethylation of astrocytic gene promoters [23] (Figure 3). We confirmed that Notch ligands are indeed expressed in committed neuronal precursor cells and young neurons, and that these ligands activate Notch signaling in the residual NSCs. Moreover, forced expression of Notch intracellular domain (NICD), which can mimic Notch signal activation in mid-gestational NSCs, upregulated NFIA expression. This in turn accelerated the demethylation of astrocytic

Figure 3



Epigenetic modifiers of astrocytic genes in NSCs. (Left) The *gfap* gene promoter in NSCs is highly methylated at early-gestational and mid-gestational stages, and STAT3 binding to the promoter is thus impeded. ESET also associates with the *gfap* promoter, inducing hypermethylation of H3K9 to repress *gfap* transcription. (Right) When Notch is activated, its intracellular domain NICD is cleaved and then translocated into the nucleus, where it forms a complex with RBP-J/k. This complex activates transcription of *Nfia*, and NFIA in turn accelerates demethylation of the *gfap* promoter by preventing the association of DNMT1. Hypoxia activates HIF1α, which associates with NICD to enhance the transcriptional activity of NICD. COUP-TFI and II also contribute to the demethylation of the *gfap* promoter. The ER stress-transducing protein OASIS induces expression of the transcription factor Gcm1, which may cause active demethylation of *gfap*. Histone H3 acetylation by RA, via its cognate nuclear receptors RAR/RXR, enables STAT3 to gain more efficient access to the *gfap* promoter. ESET is downregulated as gestation proceeds.

gene promoters by preventing their association with DNA methyltransferase 1 (DNMT1), which is essential for the maintenance of methylation patterns on the genome during cell division, and thus allowed precocious astrocyte differentiation of NSCs in response to LIF stimulation [23]. Recently, we have shown that the oxygen sensor hypoxia-inducible factor 1α (HIF1α) is important for astrocytic gene demethylation in developing NSCs by enhancing the activity of the Notch signaling pathway [24•] (Figure 3). It is generally known that embryonic tissues including the brain are hypoxic, and we further showed that astrocyte differentiation of NSCs was impaired when the embryos were incubated in a hyperoxic environment. The oxygen level surrounding NSCs is thus apparently critical for the appropriate scheduling of astrocytogenesis through its capacity to affect epigenetic modification.

The murine homologs of chicken ovalbumin upstream promoter transcription factors I and II (COUP-TFI/II) also contribute to the demethylation of the *gfap* promoter during development. Using a mouse embryonic stem cell (mESC)-derived NSC culture that recapitulates mouse central nervous system development observed *in vivo* [25,26], Naka et al. reported that in *Coup-tfI/II* double-knockdown NSCs, the hypermethylated status of the *gfap* promoter was maintained, and the switch from neurogenesis to astrocytogenesis was thereby inhibited [26].

Furthermore, knockdown of *Coup-tfI/II* in the developing mouse forebrain also resulted in impaired astrocyte differentiation of NSCs [26]. Although the mechanism remains unknown, these results indicate that COUP-TFI and II are important factors for *gfap* promoter demethylation (Figure 3).

Recently, the unfolded protein response (UPR), which is triggered by unfolded protein accumulation-induced endoplasmic reticulum (ER) stress, has been implicated in astrocyte differentiation during brain development. The ER stress transducer protein Old Astrocyte Specifically Induced Substance (OASIS) plays a key role in astrocytogenesis from NSCs by regulating *gfap* gene promoter methylation [27•] (Figure 3). OASIS belongs to the CREB/ATF family of transcription factors and modulates cell-specific or tissue-specific unfolded protein response signaling [28]. *Oasis* knockout in NSCs resulted in inhibited astrocyte differentiation, attributable to diminished DNA demethylation of the *gfap* promoter. Furthermore, expression of the transcription factor Gcm1, a mammalian homolog of Drosophila GCM [29], which is essential for glial differentiation in Drosophila, was upregulated by OASIS, and Gcm1 overexpression accelerated demethylation of the *gfap* promoter [27•] (Figure 3). Hitoshi et al. showed recently that GCM1 and GCM2 are involved in the demethylation of *Hes5* promoter DNA during the production of early-gestational NSCs from primitive neuroepithelium in mouse embryos [30••]. This report interestingly suggests that, unlike the case of NFIA on the *gfap* promoter, demethylation of the *Hes5* promoter by GCM1 and GCM2 is an active demethylation, which is independent of DNA replication. Thus, the active demethylation of astrocytic genes by GCM1 may also contribute to the acquisition of astrocytogenetic potential by NSCs during development.

Chromatin modification is also important for the regulation of astrocytogenesis from NSCs during development. The best-characterized chromatin modifications are acetylation and methylation of lysine (K) residues of the core histones, H3 and H4. An increase in histone acetylation by histone acetyltransferases causes remodeling of chromatin from a tightly to a loosely packed configuration (euchromatin), which leads to transcriptional activation. Our group suggested that histone H3 acetylation by retinoic acid (RA) via its cognate nuclear receptors, RAR/RXR, on the *gfap* promoter facilitates LIF-induced astrocyte differentiation of NSCs by allowing STAT3 to gain more efficient access to the promoter [31] (Figure 3). In contrast to acetylation, methylation of histones can result in either activation or repression of gene transcription, depending on which residue is methylated [32]. For instance, H3K4 methylation marks transcriptionally active chromatin, whereas methylated H3K9 and H3K27 mark transcriptionally inactive chromatin. The H3K9 methyltransferase ESET (also called Setdb1 or KMT1E) is known to repress gene expression in euchromatin by interacting with the co-repressor KAP1 (Trim28) [33]. Tan et al. [34•] showed that ESET is highly expressed at early stages of mouse brain development, but is downregulated as gestation proceeds. They further suggested that inactivation of *ESET* impaired neurogenesis at early embryonic stages, but accelerated astrocyte production *in vivo*. Tan et al. also noted that ESET binds to the *gfap* promoter to induce H3K9 methylation in wild-type NSCs, whereas in *ESET* mutant NSCs H3K9 is hypomethylated, suggesting that H3K9 methylation is important for the appropriate timing of astrocytogenesis in the embryonic mouse brain [34•] (Figure 3).

Closing remarks

Space constraints prevent us from introducing many other transcriptional factors that participate, directly or indirectly, in astrocyte differentiation of NSCs, notably Olig2 [35], serum response factor [36], RP58 [37••], and LIM-homeodomain transcription factor 2 (Lhx2) [38••]. The findings described here and in previous reports reveal clearly that an exquisite interplay among a great variety of factors — including transcriptional factors, epigenetic modifiers, and the environment surrounding the cells — systematically coordinates the generation of astrocytes from NSCs during brain development. Given the astrocyte's roles in diverse brain functions, the impairment of astrocyte development may contribute to neurological disorders at postnatal and adult stages, such as brain tumors, epilepsy and psychiatric diseases. For example, Alexander disease, which is caused by a mutation in GFAP, is characterized by macrocephaly, abnormal white matter, and developmental delay [39]. It has been reported that GFAP expression is increased in some Alexander disease brains, and that overexpression of wild-type GFAP in mice resulted in the formation of Rosenthal fibers, pathological marks of Alexander disease, that are indistinguishable from those found in afflicted patients [40]. In addition, gene expression profiling of mouse brains has revealed the misexpression of many genes involved in glutathione metabolism, peroxide detoxification and iron homeostasis in neurons [41]. Thus, further progress in the investigation of astrocyte development will contribute greatly to our understanding of human neurological disorders.

Acknowledgements

We thank our laboratory members for useful discussions on this topic, and Ian Smith for critical reading of the manuscript. We have been supported by a Grant-in-Aid for Scientific Research on Innovative Areas 'Neural Diversity and Neocortical Organization', and the NAIST Global COE Program (Frontier Biosciences: Strategies for Survival and Adaptation in a Changing Global Environment), from the Ministry of Education, Culture, Sports, Science and Technology.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Azevedo FA, Carvalho LR, Grinberg LT, Farfel JM, Ferretti RE, Leite RE, Jacob Filho W, Lent R, Herculano-Houzel S: **Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain**. *J Comp Neurol* 2009, **513**:532-541.

2. Marin-Padilla M: **Prenatal development of fibrous (white matter), protoplasmic (gray matter), and layer I astrocytes in the human cerebral cortex: a Golgi study**. *J Comp Neurol* 1995, **357**:554-572.

3. Tramontin AD, Garcia-Verdugo JM, Lim DA, Alvarez-Buylla A: **Postnatal development of radial glia and the ventricular zone (VZ): a continuum of the neural stem cell compartment**. *Cereb Cortex* 2003, **13**:580-587.

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

5. Juliandi B, Abematsu M, Nakashima K: **Epigenetic regulation in neural stem cell differentiation**. *Dev Growth Differ* 2010, **52**:493-504.

6. • Ge WP, Miyawaki A, Gage FH, Jan YN, Jan LY: **Local generation of glia is a major astrocyte source in postnatal cortex**. *Nature* 2012, **484**:376-380.

This paper shows that the majority of astrocytes in the postnatal cortex are produced by local proliferation of differentiated cells during the early postnatal period.

7. Ernst M, Jenkins BJ: **Acquiring signalling specificity from the cytokine receptor gp130**. *Trends Genet* 2004, **20**:23-32.

8. Bonni A, Sun Y, Nadal-Vicens M, Bhatt A, Frank DA, Rozovsky I, Stahl N, Yancopoulos GD, Greenberg ME: **Regulation of gliogenesis in the central nervous system by the JAK–STAT signaling pathway**. *Science* 1997, **278**:477-483.

9. Rajan P, McKay RD: **Multiple routes to astrocytic differentiation in the CNS**. *J Neurosci* 1998, **18**:3620-3629.

10. Nakashima K, Yanagisawa M, Arakawa H, Kimura N, Hisatsune T, Kawabata M, Miyazono K, Taga T: **Synergistic signaling in fetal brain by STAT3–Smad1 complex bridged by p300**. *Science* 1999, **284**:479-482.

11. Bugga L, Gadient RA, Kwan K, Stewart CL, Patterson PH: **Analysis of neuronal and glial phenotypes in brains of mice deficient in leukemia inhibitory factor**. *J Neurobiol* 1998, **36**:509-524.

12. Koblar SA, Turnley AM, Classon BJ, Reid KL, Ware CB, Cheema SS, Murphy M, Bartlett PF: **Neural precursor differentiation into astrocytes requires signaling through the leukemia inhibitory factor receptor**. *Proc Natl Acad Sci U S A* 1998, **95**:3178-3181.

13. Nakashima K, Wiese S, Yanagisawa M, Arakawa H, Kimura N, Hisatsune T, Yoshida K, Kishimoto T, Sendtner M, Taga T: **Developmental requirement of gp130 signaling in neuronal survival and astrocyte differentiation**. *J Neurosci* 1999, **19**:5429-5434.

14. He F, Ge W, Martinowich K, Becker-Catania S, Coskun V, Zhu W, Wu H, Castro D, Guillemot F, Fan G *et al.*: **A positive autoregulatory loop of Jak–STAT signaling controls the onset of astrogliogenesis**. *Nat Neurosci* 2005, **8**:616-625.

15. Barnabe-Heider F, Wasylnka JA, Fernandes KJ, Porsche C, Sendtner M, Kaplan DR, Miller FD: **Evidence that embryonic neurons regulate the onset of cortical gliogenesis via cardiotrophin-1**. *Neuron* 2005, **48**:253-265.

16. Sun Y, Nadal-Vicens M, Misono S, Lin MZ, Zubiaga A, Hua X, Fan G, Greenberg ME: **Neurogenin promotes neurogenesis and inhibits glial differentiation by independent mechanisms**. *Cell* 2001, **104**:365-376.

17. Guillemot F: **Spatial and temporal specification of neural fates by transcription factor codes**. *Development* 2007, **134**:3771-3780.

18. Hirabayashi Y, Suzki N, Tsuboi M, Endo TA, Toyoda T, Shinga J, Koseki H, Vidal M, Gotoh Y: **Polycomb limits the neurogenic competence of neural precursor cells to promote astrogenic fate transition**. *Neuron* 2009, **63**:600-613.

19. • Li X, Newbern JM, Wu Y, Morgan-Smith M, Zhong J, Charron J, Snider WD: **MEK is a key regulator of gliogenesis in the developing brain**. *Neuron* 2012, **75**:1035-1050.

This study shows that RAF/MEK/ERK pathway plays an important role in astrocytogenesis from NSCs through its regulation of the JAK/STAT3 pathway.

20. Cahoy JD, Emery B, Kaushal A, Foo LC, Zamanian JL, Christopherson KS, Xing Y, Lubischer JL, Krieg PA, Krupenko SA *et al.*: **A transcriptome database for astrocytes, neurons, and oligodendrocytes: a new resource for understanding brain development and function**. *J Neurosci* 2008, **28**:264-278.

21. Takizawa T, Nakashima K, Namihira M, Ochiai W, Uemura A, Yanagisawa M, Fujita N, Nakao M, Taga T: **DNA methylation is a critical cell-intrinsic determinant of astrocyte differentiation in the fetal brain**. *Dev Cell* 2001, **1**:749-758.

22. Namihira M, Nakashima K, Taga T: **Developmental stage dependent regulation of DNA methylation and chromatin modification in a immature astrocyte specific gene promoter**. *FEBS Lett* 2004, **572**:184-188.

23. Namihira M, Kohyama J, Semi K, Sanosaka T, Deneen B, Taga T, Nakashima K: **Committed neuronal precursors confer astrocytic potential on residual neural precursor cells**. *Dev Cell* 2009, **16**:245-255.

24. • Mutoh T, Sanosaka T, Ito K, Nakashima K: **Oxygen levels epigenetically regulate fate switching of neural precursor cells via hypoxia-inducible factor 1alpha-notch signal interaction in the developing brain**. *Stem Cells* 2012, **30**:561-569.

This report indicates that in the environment surrounding NSCs, a low oxygen level is important for astrocytic gene demethylation in developing NSCs by enhancing the activity of the Notch signaling pathway.

25. Okada Y, Matsumoto A, Shimazaki T, Enoki R, Koizumi A, Ishii S, Itoyama Y, Sobue G, Okano H: **Spatiotemporal recapitulation of central nervous system development by murine embryonic stem cell-derived neural stem/progenitor cells**. *Stem Cells* 2008, **26**:3086-3098.

26. Naka H, Nakamura S, Shimazaki T, Okano H: **Requirement for COUP-TFI and II in the temporal specification of neural stem cells in CNS development**. *Nat Neurosci* 2008, **11**:1014-1023.

27. • Saito A, Kanemoto S, Kawasaki N, Asada R, Iwamoto H, Oki M, Miyagi H, Izumi S, Sanosaka T, Nakashima K, Imaizumi K: **Unfolded protein response, activated by OASIS family transcription factors, promotes astrocyte differentiation**. *Nat Commun* 2012, **3**:967.

The authors show that the ER stress transducer protein OASIS plays a key role in astrocytogenesis of NSCs via demethylation of the gfap gene promoter, and that Gcm1, the downstream target of OASIS, may achieve this demethylation.

28. Kondo S, Murakami T, Tatsumi K, Ogata M, Kanemoto S, Otori K, Iseki K, Wanaka A, Imaizumi K: **OASIS, a CREB/ATF-family member, modulates UPR signalling in astrocytes**. *Nat Cell Biol* 2005, **7**:186-194.

29. Kim J, Jones BW, Zock C, Chen Z, Wang H, Goodman CS, Anderson DJ: **Isolation and characterization of mammalian homologs of the *Drosophila* gene glial cells missing**. *Proc Natl Acad Sci U S A* 1998, **95**:12364-12369.

30. •• Hitoshi S, Ishino Y, Kumar A, Jasmine S, Tanaka KF, Kondo T, Kato S, Hosoya T, Hotta Y, Ikenaka K: **Mammalian Gcm genes induce Hes5 expression by active DNA demethylation and induce neural stem cells**. *Nat Neurosci* 2011, **14**:957-964.

This report shows that Gcm1 and Gcm2 are involved in demethylation of the Hes5 promoter during the production of early-gestational NSCs from primitive neuroepithelium in mouse embryos, and that demethylation of Hes5 promoter by GCM1 and GCM2 is the active demethylation independently of DNA replication.

31. Asano H, Aonuma M, Sanosaka T, Kohyama J, Namihira M, Nakashima K: **Astrocyte differentiation of neural precursor cells is enhanced by retinoic acid through a change in epigenetic modification**. *Stem Cells* 2009, **27**:2744-2752.

32. Yoo CB, Jones PA: **Epigenetic therapy of cancer: past, present and future**. *Nat Rev Drug Discov* 2006, **5**:37-50.

33. Schultz DC, Ayyanathan K, Negorev D, Maul GG, Rauscher FJ 3rd: **SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins**. *Genes Dev* 2002, **16**:919-932.

34. • Tan SL, Nishi M, Ohtsuka T, Matsui T, Takemoto K, Kamio-Miura A, Aburatani H, Shinkai Y, Kageyama R: **Essential roles of the histone methyltransferase ESET in the epigenetic control of neural progenitor cells during development**. *Development* 2012, **139**:3806-3816.
This study indicates that H3K9 methylation by ESET is important for the appropriate timing of astrocytogenesis in the embryonic mouse brain.

35. Cai J, Chen Y, Cai WH, Hurlock EC, Wu H, Kernie SG, Parada LF, Lu QR: **A crucial role for Olig2 in white matter astrocyte development**. *Development* 2007, **134**:1887-1899.

36. Lu PP, Ramanan N: **A critical cell-intrinsic role for serum response factor in glial specification in the CNS**. *J Neurosci* 2012, **32**:8012-8023.

37. •• Hirai S, Miwa A, Ohtaka-Maruyama C, Kasai M, Okabe S, Hata Y, Okado H: **RP58 controls neuron and astrocyte differentiation by downregulating the expression of Id1–4 genes in the developing cortex**. *EMBO J* 2012, **31**:1190-1202.
This report indicates that the transcription factor RP58 downregulates all four Id genes in the developing cerebral cortex. RP58 KO mice display enhanced astrocytogenesis, and the phenotypes of their NSCs were mimicked by the overexpression of all Id genes in wild-type cortical progenitors. These observations implicate RP58 as a novel key regulator that controls the differentiation of NSCs and restricts astrocytogenesis by repressing Id genes during corticogenesis.

38. •• Subramanian L, Sarkar A, Shetty AS, Muralidharan B, Padmanabhan H, Piper M, Monuki ES, Bach I, Gronostajski RM, Richards LJ *et al.*: **Transcription factor Lhx2 is necessary and sufficient to suppress astrogliogenesis and promote neurogenesis in the developing hippocampus**. *Proc Natl Acad Sci U S A* 2011, **108**:E265-E274.
Subramanian et al. show that Lhx2 KO in the embryonic hippocampus by in utero electroporation, and in organotypic slice culture, caused the premature production of astrocytes at stages when neurons are normally generated. Furthermore, Lhx2 overexpression was able to override and suppress the activation of the gfap promoter by NFIA. However, this activity of Lhx2 is spatially restricted to the hippocampus: loss of Lhx2 function in the neocortex did not result in premature astrogliogenesis. This study therefore suggests that Lhx2 plays an important role in the neuron–glia cell fate decision in the hippocampus.

39. Quinlan RA, Brenner M, Goldman JE, Messing A: **GFAP and its role in Alexander disease**. *Exp Cell Res* 2007, **313**:2077-2087.

40. Messing A, Head MW, Galles K, Galbreath EJ, Goldman JE, Brenner M: **Fatal encephalopathy with astrocyte inclusions in GFAP transgenic mice**. *Am J Pathol* 1998, **152**:391-398.

41. Hagemann TL, Gaeta SA, Smith MA, Johnson DA, Johnson JA, Messing A: **Gene expression analysis in mice with elevated glial fibrillary acidic protein and Rosenthal fibers reveals a stress response followed by glial activation and neuronal dysfunction**. *Hum Mol Genet* 2005, **14**:2443-2458.

Uesaka et al. BMC Genomics 2014, 15:35
http://www.biomedcentral.com/1471-2164/15/35

BMC Genomics

RESEARCH ARTICLE Open Access

Bidirectional promoters are the major source of gene activation-associated non-coding RNAs in mammals

Masahiro Uesaka[1,2], Osamu Nishimura[1,3], Yasuhiro Go[4,5], Kinichi Nakashima[2], Kiyokazu Agata[1] and Takuya Imamura[1,2*]

Abstract

Background: The majority of non-coding RNAs (ncRNAs) involved in mRNA metabolism in mammals have been believed to downregulate the corresponding mRNA expression level in a pre- or post-transcriptional manner by forming short or long ncRNA-mRNA duplex structures. Information on non-duplex-forming long ncRNAs is now also rapidly accumulating. To examine the directional properties of transcription at the whole-genome level, we performed directional RNA-seq analysis of mouse and chimpanzee tissue samples.

Results: We found that there is only about 1% of the genome where both the top and bottom strands are utilized for transcription, suggesting that RNA-RNA duplexes are not abundantly formed. Focusing on transcription start sites (TSSs) of protein-coding genes revealed that a significant fraction of them contain switching-points that separate antisense- and sense-biased transcription, suggesting that head-to-head transcription is more prevalent than previously thought. More than 90% of head-to-head type promoters contain CpG islands. Moreover, CCG and CGG repeats are significantly enriched in the upstream regions and downstream regions, respectively, of TSSs located in head-to-head type promoters. Genes with tissue-specific promoter-associated ncRNAs (pancRNAs) show a positive correlation between the expression of their pancRNA and mRNA, which is in accord with the proposed role of pancRNA in facultative gene activation, whereas genes with constitutive expression generally lack pancRNAs.

Conclusions: We propose that single-stranded ncRNA resulting from head-to-head transcription at GC-rich sequences regulates tissue-specific gene expression.

Keywords: Bidirectional promoter, Non-coding RNA, CpG island, Directional RNA-Seq, Gene activation

Background

Protein-coding regions account for only about 1.5% of the human genome [1], but the FANTOM Consortium and the ENCODE Project Consortium revealed that more than 62% of the genomic DNA acts as a template for transcription [2,3], indicating that there are a large number of non-coding RNAs (ncRNAs) in living cells. Recently, many functional ncRNAs have been identified. It is well known that small RNAs, such as miRNAs and piRNAs, act in post-transcriptional regulation by forming RNA-RNA duplexes [4,5]. In addition to these RNAs, many kinds of long ncRNAs have been shown to function in post-transcriptional regulation, such as RNA editing, splicing and translation, by forming RNA-RNA duplexes [6-13]. Indeed, 4,520 sense-antisense transcript (SAT) pairs in mice have the potential to form RNA-RNA duplexes [14]. RNA-RNA duplexes also play a role in transcriptional gene silencing through DNA methylation and histone modifications [15-18]. Thus, it is clear that the formation of RNA-RNA duplexes is important for the mRNA silencing triggered by ncRNA.

However, several studies have reported that some long ncRNAs cause transcriptional activation of genes without

* Correspondence: imamura@scb.med.kyushu-u.ac.jp
[1]Department of Biophysics and Global COE Program, Graduate School of Science, Kyoto University, Kitashirakawa-Oiwake, Sakyo-ku, Kyoto 606-8502, Japan
[2]Department of Stem Cell Biology and Medicine, Graduate School of Medical Sciences, Kyushu University, 3-1-1 Maidashi, Higashi-ku, Fukuoka 812-8581, Japan
Full list of author information is available at the end of the article

forming RNA-RNA duplexes. For example, *HOTTIP*, a long intergenic ncRNA (lncRNA) transcribed from the 5'-end of the *HOXA* locus, binds to an adaptor protein, WD repeat-containing protein 5 (WDR5), which in turn recruits the mixed-lineage leukaemia (MLL) histone methyltransferase complex [19]. With the help of *HOTTIP*-WDR5-MLL1 interaction, several distantly located target genes are brought into close contact through tertiary structure formation, resulting in trimethylation of histone H3K4 and gene activation. Moreover, a recent study showed that *DBE-T*, a chromatin-associated ncRNA, is selectively transcribed from the chromosome 4q35 region in facioscapulohumeral muscular dystrophy patients and coordinates the transcription of 4q35 genes [20]. *DBE-T* recruits the Trithorax group protein ASH1L, a histone-lysine N-methyltransferase, to the DNA template for *DBE-T*, driving histone H3K36 dimethylation and 4q35 gene transcription. Therefore, lncRNAs acting together with chromosomal proteins are thought to regulate gene functions in an RNA-RNA hybridization-independent manner. However, in contrast to small RNAs, there are few reports about the functional properties of single-strand ncRNAs that act without forming RNA-RNA duplexes.

In mammals, CpG islands (CGIs) in promoter regions tend to show bidirectional promoter activity [21,22]. CGIs are utilized for bidirectional transcription in a head-to-head (HtH) manner. Our previous reports have shown that, in contradiction to the prevalent idea that ncRNAs other than classical ncRNAs (tRNA, rRNA, snRNA and snoRNA) downregulate target gene expression, antisense long ncRNAs derived from promoter regions of their respective protein-coding genes activate the expression of those genes via sequence-specific DNA demethylation [23,24]. We termed these antisense long (>200 nt) ncRNAs "promoter-associated ncRNAs" (pancRNAs). At present, little is known about the concerted expression of mRNAs and antisense transcripts produced in their 5'-flanking regions, and comprehensive transcriptome analysis focusing on the bidirectional transcription of mRNA and pancRNA has not been performed. We do not yet know the sequence characteristics of bidirectionally transcribed promoter regions. Here, we examine whether there is a correlation between the expression of sense and antisense transcripts at the genome-wide level using directional RNA-seq. We map the origin of the sense and antisense transcripts found by directional RNA-seq to determine the prevalence of HtH transcript pairs from CGI promoters. We propose that highly expressed antisense transcripts derived from bidirectional transcription start sites (TSSs) show coordinated transcription with the corresponding protein-coding genes.

Results

Both top and bottom strands are utilized in a small fraction of the genome

We analyzed directional RNA-seq data in order to distinguish sense and antisense transcripts in the mouse cerebral cortex, cerebellum and heart, and in the chimpanzee cerebral cortex and cerebellum [DDBJ: DRA000860]. On average, we obtained 76.3 ± 1.3 million and 228.3 ± 10.7 million reads per sample from the first and second runs of Illumina HiSeq 2000, respectively (Additional file 1: Table S1). The average number of reads passing the read trimming was 72.1 ± 1.2 million and 198.3 ± 9.1 million for the first and second runs, respectively. We mapped the valid reads onto the reference genome sequences using TopHat (see Methods). We used the human instead of the chimpanzee genome as a reference for the chimpanzee reads (see Discussion for the reason). The average percentage of uniquely mapped reads in the valid reads was 86.1% for the mouse cerebral cortex, 85.4% for the mouse cerebellum, 72.6% for the mouse heart, 78.1% for the chimpanzee cerebral cortex and 82.0% for the chimpanzee cerebellum (Additional file 1: Table S1, S2). After we removed duplicate sequences, the average number of uniquely mapped reads in two replicates of each tissue sample was 19.2 million reads for the mouse cerebral cortex, 30.3 million reads for the mouse cerebellum, 18.1 million reads for the mouse heart, 19.0 million reads for the chimpanzee cerebral cortex and 22.5 million reads for the chimpanzee cerebellum. Removal of duplicated reads smoothed the unexpected protruding clusters of reads, possibly derived from PCR bias during library preparation, as shown in Additional file 2: Figure S1. The average ratio of top strand-mapped reads to bottom strand-mapped reads in a sample was 1.0 (Additional file 2: Figure S2 and Additional file 1: Table S1). We confirmed that candidate pancRNAs at *Pacsin1* and *Kcnmb4* (*pancPacsin1* and *pancKcnmb4*) were transcribed from the opposite DNA strand compared to their mRNAs, as expected, using strand-specific RT-PCR (Figure 1A, B). Although we did not confirm the functionality of the candidate pancRNAs, we refer to these transcripts as a fraction of ncRNAs based on their lower coding potential as explained later (Additional file 2: Figure S5A, B). The results of the RT-PCR analysis support the validity of our directional RNA-seq analyses. Then, we calculated the reads per kilobase per million mapped reads (RPKM) of protein-coding genes in the two replicates in order to confirm the reproducibility of our analysis. The Kendall's tau correlation between the two replicates of each tissue sample was > 0.96 ($p < 2.2e^{-16}$). Therefore, we merged the data from these two replicates for all samples and used them for the following analyses.

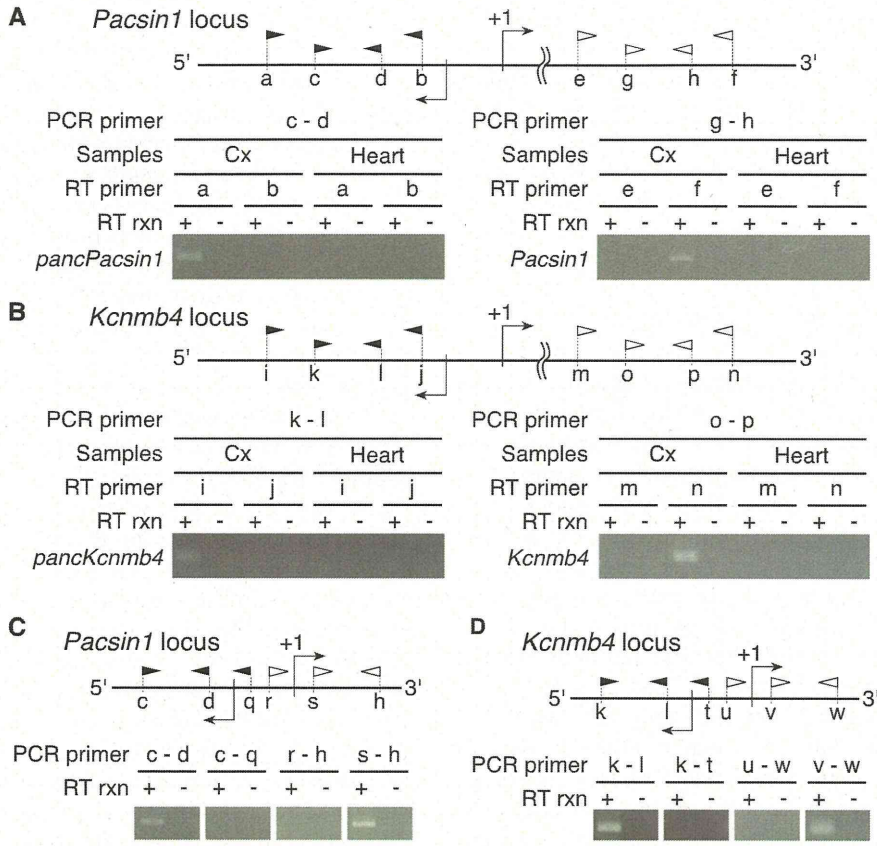


Figure 1 Determination of the transcriptional direction of pancRNAs and mRNAs. At the top of each panel, there is a schematic representation of promoter regions of *Pacsin1* **(A)(C)** and *Kcnmb4* **(B)(D)**. Filled and open arrowheads represent the primer sets used for the strand-specific RT-PCR analysis for expression of pancRNAs and mRNAs, respectively. Detailed primer information is given in Additional file 1: Table S9. **(A)(B)** Strand-specific RT-PCR analysis for expression of mRNA and pancRNAs in the mouse cerebral cortex (Cx) and heart (Heart). **(C)(D)** RT-PCR analysis for expression of mRNA and pancRNAs in the mouse cerebral cortex. + and - mean the use and lack of reverse transcriptase in the reverse transcription (RT rxn).

In the cerebral cortex, cerebellum and heart, the transcribed regions for polyA+ RNA were found to account for 25.0%, 30.0% and 21.6% of the mouse genome, respectively (Table 1). Next, we examined how many genomic regions were utilized for both sense and antisense transcription. Overlapping transcription for polyA+ RNAs was found in only 0.7%, 1.3% and 0.7% of the mouse genome in the cerebral cortex, cerebellum and heart, respectively (Table 1). A similar transcriptional landscape was found when chimpanzee samples were analyzed (Additional file 1: Table S3).

We calculated the ratio of top strand-mapped reads to bottom strand-mapped reads in the bidirectionally transcribed regions (Additional file 2: Figure S3A, B, C). The results showed that, even if the regions are bidirectionally transcribed, most of the regions show a biased expression pattern in terms of directional transcription. The mapping information in the bidirectionally transcribed regions was subgrouped into top strand, bottom strand and intergenic regions of mouse genes (Additional file 2: Figure S3F, G, H). Significantly large fractions of the top strand- and bottom strand-mapped

Table 1 The percentage of transcribed regions in the whole genome

	Transcribed regions	Unidirectionally transcribed regions[a]	Bidirectionally transcribed regions[b]
Cerebral cortex	25.0%	24.3%	0.7%
Cerebellum	30.0%	28.7%	1.3%
Heart	21.6%	20.9%	0.7%

[a]Regions where either sense or antisense transcripts (but not both) originated.
[b]Regions where both sense and antisense transcripts originated.

reads were thereby confirmed to be associated with the top and bottom strands of mouse genes, respectively. In intergenic regions, we also found biased transcription in terms of the directionality. Similarly, biased transcription was also found when chimpanzee samples were analyzed (Additional file 2: Figure S3D, E, I, J).

Taken together, these data showed that strand bias of transcription occurred on a genome-wide level. Either top or bottom strand was preferentially utilized depending on the tissue.

Genome-wide production of ncRNAs that do not form RNA-RNA duplexes

Our previous studies demonstrated that antisense transcripts from promoter regions could activate the sense transcription of the same locus [23,24]. Hence, we analyzed HtH transcript pairs, rather than overlapping transcription. In order to examine the switching-point of the bi-transcriptional direction, we focused on the genomic regions around TSSs of the reference genes. First, we adjusted the TSS of each reference gene according to the mapped reads of each tissue sample (see Methods). This adjustment is important for determining the precise distribution of mapped reads around TSSs. In fact, the ENCODE project showed that approximately 48% of the CAGE-identified TSSs are located hundreds of base pairs away from annotated GENCODE TSSs, indicating the requirement for this adjustment of TSSs [25]. Then, we examined the distribution of sense and antisense mapped reads around the TSS of each mouse protein-coding gene (Figure 2A and Additional file 2: Figure S2, S4A, E). In order to focus on the ncRNA-expressing promoters, we removed the HtH-type promoters driving protein-coding gene expression in both directions from our datasets. We examined the longest open reading frame (ORF) in each region between +1 and +1,000 bp and those between −1000 and −1 bp relative to the TSS, respectively. The mean length of the longest ORFs in the upstream and downstream regions is 191.5 and 319.6, respectively, in the mouse dataset. In the chimpanzee dataset, the mean length of the longest ORFs in the upstream and downstream regions is 213.0 and 305.0, respectively. Next, we examined the distribution of the longest ORF size in the mouse and chimpanzee dataset (Additional file 2: Figure S5A, B). There is one peak around 200 nt for the upstream region. On the other hand, there are two peaks around 200 and 900 nt for the downstream region. The 900-nt-peak seemed to reflect the fraction consisting of protein-coding genes. Moreover, we examined whether regions between −1,000 and −1 bp relative to the TSS contained any conserved protein domains by using NCBI's Conserved Domain Database [26]. Only 1.9% and 4.8% of all regions from −1 to −1,000 bp relative to the TSS contain any conserved protein domains in the mouse and chimpanzee dataset, respectively. In contrast, 20.3% and 15.5% of all regions from +1 to +1,000 bp relative to the TSS contain conserved protein domains in the mouse and chimpanzee dataset, respectively. These results suggest that the vast majority of the upstream regions in our datasets produced ncRNAs, although we cannot completely exclude the possibility that a fraction of these antisense transcripts encode very short proteins.

In order to investigate if antisense transcription occurs in conjunction with transcription of the corresponding mRNA, we examined the distribution of sense and antisense mapped reads around the TSSs of mouse genes. Toward this end, we selected the 100 genes with the most-highly expressed pancRNAs and the 100 genes with the most-weakly expressed pancRNAs, as indicated by RPKM (Figure 2B and Additional file 2: Figure S4B, F). In this selection, we did not consider mRNA expression level for the selection of genes. For the RPKM calculation of the pancRNAs, only antisense mapped reads in the upstream region of protein-coding genes were counted. For estimation of the promoter activity of protein-coding genes, we focused on the region between +1 and +1,000 bp relative to the TSSs. Both in the mouse and chimpanzee samples, more sense reads were mapped to the protein-coding genes with pancRNAs showing the top 100 ranked RPKM than to those with pancRNAs showing the bottom 100 ranked RPKM ($p < 0.001$; Additional file 2: Figure S4J, N, S6A, Table 2, and Additional file 1: Table S4). When we calculated RPKM of the protein-coding genes based on the reference gene structure, we again found that the protein-coding genes with pancRNAs showing the top 100 ranked RPKM were more highly expressed than those with pancRNAs showing the bottom 100 ranked RPKM ($p < 0.001$; Additional file 2: Figure S6B). From the 100 regions with the most-highly expressed pancRNAs, we extracted three types of genomic regions: 1) the expression level of antisense transcript from the upstream region of the TSS is at least five times higher than that from the downstream region, 2) the expression level of antisense transcript from the upstream region of the TSS is at least two times lower than that from the downstream region, and 3) the remaining regions not meeting condition 1) or 2). Then, we examined the distribution of sense and antisense mapped reads in each subgroup (Figure 2C and Additional file 2: Figure S4C, G). Although we cannot rule out a possible short association between ncRNA and the corresponding mRNA at their 5'-ends, RT-PCR detection of transcripts derived from pancRNA-bearing gene loci *Pacsin1* and *Kcnmb4* supported the positive correlation between pancRNA and mRNA expression (Figure 1). We investigated two representative genes and confirmed that pancRNA and