

SCIENTIFIC REPORTS



OPEN

Identification of genes associated with the astrocyte-specific gene *Gfap* during astrocyte differentiation

Received: 23 January 2015

Accepted: 16 March 2016

Published: 04 April 2016

Kenji Ito¹, Tsukasa Sanosaka², Katsuhide Igarashi³, Maky Ideta-Otsuka³, Akira Aizawa¹, Yuichi Uosaki¹, Azumi Noguchi¹, Hirokazu Arakawa¹, Kinichi Nakashima² & Takumi Takizawa¹

Chromosomes and genes are non-randomly arranged within the mammalian cell nucleus, and gene clustering is of great significance in transcriptional regulation. However, the relevance of gene clustering and their expression during the differentiation of neural precursor cells (NPCs) into astrocytes remains unclear. We performed a genome-wide enhanced circular chromosomal conformation capture (e4C) to screen for genes associated with the astrocyte-specific gene glial fibrillary acidic protein (*Gfap*) during astrocyte differentiation. We identified 18 genes that were specifically associated with *Gfap* and expressed in NPC-derived astrocytes. Our results provide additional evidence for the functional significance of gene clustering in transcriptional regulation during NPC differentiation.

An increasing amount of evidence supports the importance of spatial organization of the genome in the nuclei of higher eukaryotes¹. Chromosomes and genes are non-randomly arranged and occupy preferential positions within the nucleus². Moreover, these arrangements are associated with gene regulation because the sub-nuclear positions of genes change along with alterations in their transcriptional states^{3–6}. For instance, in naïve CD4⁺T helper cells, there is an inter-chromosomal association between the regulatory region of the TH2 cytokine locus and the interferon γ (*Ifng*) promoter region, where both are repressed⁷. On the other hand, in erythroid cells, Klf1-regulated genes including globins preferentially associate at a limited number of transcriptional factories containing high levels of Klf1 once activated⁸. Other observations based on chromosome conformation capture (3C) and its derivative techniques (4C, 5C, ChIA-PET) have shown that gene associations play roles in transcriptional regulation^{9–12}. These techniques are essential for revealing three-dimensional information regarding the spatial proximity of DNA within the cell nucleus^{13,14}.

Neural precursor cells (NPCs) in the central nervous system can self-renew and differentiate into neurons mid-gestation, and then into astrocytes and oligodendrocytes only after late-gestation¹⁵. Differentiation of NPCs is temporally and spatially regulated by several factors including cytokines and epigenetic modifications^{16,17}. NPCs from mouse telencephalon at late gestation (e.g., embryonic day [E] 14.5) are competent to differentiate into astrocytes upon stimulation with leukemia inhibitory factor (LIF)^{18,19}. LIF activates the transcription factor STAT3, which then binds to the promoter of an astrocyte specific gene, glial fibrillary acidic protein (*Gfap*), to induce its expression^{19,20}. This is of great relevance in astrocytogenesis, since mice lacking a common receptor for LIF, gp130, are largely devoid of *Gfap*-positive astrocytes. These astrocytes show a lower ability to support the survival of neurons²¹. In addition, DNA demethylation and chromatin remodeling in STAT3 binding motifs on the *Gfap* promoter are essential for *Gfap* expression²².

Gfap gene loci have been shown to undergo a shift toward a more internal location upon transcriptional activation⁶. Furthermore, genomic regions adjacent to nuclear lamina are replaced as gene expression programs

¹Department of Pediatrics, Graduate School of Medicine, Gunma University, 3-39-22 Showa-machi, Maebashi, Gunma 371-8511, Japan. ²Stem Cell Biology and Medicine, Department of Stem cell Biology and Medicine, Graduate School of Medical Sciences, Kyushu University, 3-1-1 Maidashi, Higashi-ku, Fukuoka, 812-8582, Japan. ³Life Science Tokyo Advanced Research Center (L-StaR), Pharmacy and Pharmaceutical Science, Hoshi University, 2-4-41 Ebara, Shinagawa-ku, Tokyo 142-5801, Japan. Correspondence and requests for materials should be addressed to T.T. (email: takizawt@gunma-u.ac.jp)

change during astrocyte differentiation from NPCs²³. This indicates robust conversion of genome localization during astrocytogenesis; however, little is known about the relevance of gene clustering in NPC differentiation.

In this study, we screened for genes that associate with *Gfap* during the astrocyte differentiation of NPCs by using enhanced circular chromosome conformation capture with minor modifications (modified e4C). We looked for a correlation between gene clustering and transcriptional activities by comparing data from modified e4C and expression arrays. We identified 18 genes associated with *Gfap* that are also expressed specifically in LIF-induced astrocytes. DNA fluorescence *in situ* hybridization (FISH) confirmed the clustering of some genes and *Gfap*. These findings support the possibility that the association of co-expressing genes is involved in astrocyte differentiation.

Results

Genome-wide screening of genes specifically associated with *Gfap* and expressed in NPC-derived astrocytes.

As a first step toward identifying genes clustered with and regulated similarly to *Gfap* during astrocyte differentiation, we decided to perform a modified e4C assay with a few modifications⁸. NPCs derived from E14.5 mouse brains can differentiate into astrocytes after being cultured *in vitro* for more than 4 days in the presence of the astrocyte-inducing cytokine LIF¹⁹. We isolated neuroepithelial cells from the telencephalon of E14.5 mice and cultured them for 5 consecutive days (designated as NPCs). After one passage, the NPCs were further cultured for 4 days with LIF to differentiate them into astrocytes (designated as LIF+ cells) (Fig. 1A). As reported previously, under these conditions, ~20% of NPCs differentiate into astrocytes as judged by immunofluorescence labeling of the astrocyte marker *GFAP* (Fig. 1B)^{6,19}. The NPCs grown in extended culture without LIF (LIF- cells) were also tested as a control (Fig. 1A,B).

As “bait” for the e4C assay, we used a genomic region containing a STAT3 cognitive sequence on the *Gfap* promoter, the *Gfap* STAT3-binding site (GSBS). The GSBS is located ~1.5 kb upstream of the transcription start site and is a prerequisite for *Gfap* transcription during astrocytogenesis from NPCs¹⁹. We first tried *Bgl*II digestion of the flanking regions of GSBS. However, the digestion efficiency in this region as assessed with quantitative PCR was much lower in cells that did not express *Gfap* than those that did express *Gfap* (20.8% vs. 61.3%). We assumed the insufficient digestion was due to highly compacted chromatin around the GSBS in those cells²². To improve accessibility of restriction enzymes to the chromatin, we added an extra step of hydrochloric acid treatment to the original e4C protocol (Fig. 1C). Indeed, this achieved comparable digestion efficiency at the GSBS region in different types of cells (Fig. 1D) and helped to identify a large number of e4C peaks in both *cis* and *trans* in two biological replicates. As expected, many peaks were found around the bait locus, and the ratio between the e4C peak number and chromosome size was the highest for chromosome 11, which encodes *Gfap* in all the types of cells tested (Fig. 1E,F).

Trans-chromosomal interactions were also found in all types of cells tested (Fig. 2A). A number of positive probes were not replicated between two biological repeats (Figure S1A); this might be partly due to a resolution difference between the 3C assay and microarray probes. The former sometimes has cytological but not necessarily molecular resolution²⁴, while the arrays represent molecular resolution and contain a probe or two for most target fragments between *Bgl*II and *Nla*III with expected lengths within 200 bp (Figure S1B). Therefore, we only focused on replicated positive probes. The analysis of distances of positive probes from the nearest genes revealed that probes 10–30 kbp away from the transcriptional start or end sites were enriched with e4C compared to all probes on the array (Figure S2). Indeed, as many as ~1,000 genes were found within 50 kbp from the positive probes that we defined as being associated with *Gfap* in each cell type (Fig. 2B and Supplemental Table S2), likely because of the relatively lower resolution of the 3C or its derivative assay²⁴. Among them, 587 genes were associated with *Gfap* exclusively in LIF+ cells (Fig. 2B). We also found considerable overlap (564 genes) between NPCs and LIF- cells but less overlap between NPCs and LIF+ cells (430 genes) or LIF+ and LIF- cells (295 genes). These results show distinct associations between *Gfap* and other genes in LIF+ cells versus NPCs and LIF- cells.

To identify genes that are simultaneously associated with *Gfap* gene loci and expressed specifically in LIF+ cells, gene expression profiles in each cell type were examined by Affymetrix GeneChip analysis. To normalize gene expression levels from different samples, we adopted the Perccellome method, which provides “per cell” read-outs in copy numbers of mRNA by adding a set of external spike mRNAs in proportion to the DNA content as a substitute for the measurement of cell numbers in the sample²⁵. We found 2,083 genes in LIF+ cells and 2,404 in LIF- cells, with more than a two-fold increase in expression in NPCs. There were 1,132 overlapping genes between LIF+ and LIF- cells and 951 genes exclusively expressed in LIF+ cells (Fig. 2C and Supplemental Table S3). By referring to genes associated in LIF+ cells in the e4C assay, we found 18 genes that were associated with *Gfap* and expressed at particularly high levels in LIF+ cells (Table 1 and Fig. 2D).

Validation of GeneChip array and modified e4C results. Among the putative 18 genes, we selected *Ogn*, *Osmr*, *Ecr4*, *A2m*, and *Gab1* for further analysis because they have been implicated as playing important roles in the central nervous system^{26–30}. To affirm the results of the GeneChip analysis, we tested the expressions of the five genes and *Gfap* by RT-qPCR. As expected, mRNA expression levels of the six genes were all significantly increased 72 and 96 h after LIF stimulation (Fig. 3).

The long-range interactions identified by 4C technology need to be verified by completely independent methods such as FISH because 4C technology shows average chromosome conformations from millions of nuclei^{31,32}. Therefore, we next performed DNA FISH for each of the five putative associating genes (Fig. 4A and Figure S3). In addition to these five genes, we also tested genes that were negative in modified e4C and located on different chromosomes, namely *Ahnak* on chromosome 19 and *Nme8* on chromosome 13. The GeneChip analysis showed that *Ahnak* is exclusively expressed in LIF+ cells, while *Nme8* is expressed at miniscule levels in NPCs, LIF+, and LIF- cells. Three genes (*Ogn*, *Osmr*, and *Ecr4*) out of the five putative genes show a significant increase in the number of cells with associated alleles in the LIF+ culture (Fig. 4B). Although not significant, similar tendencies

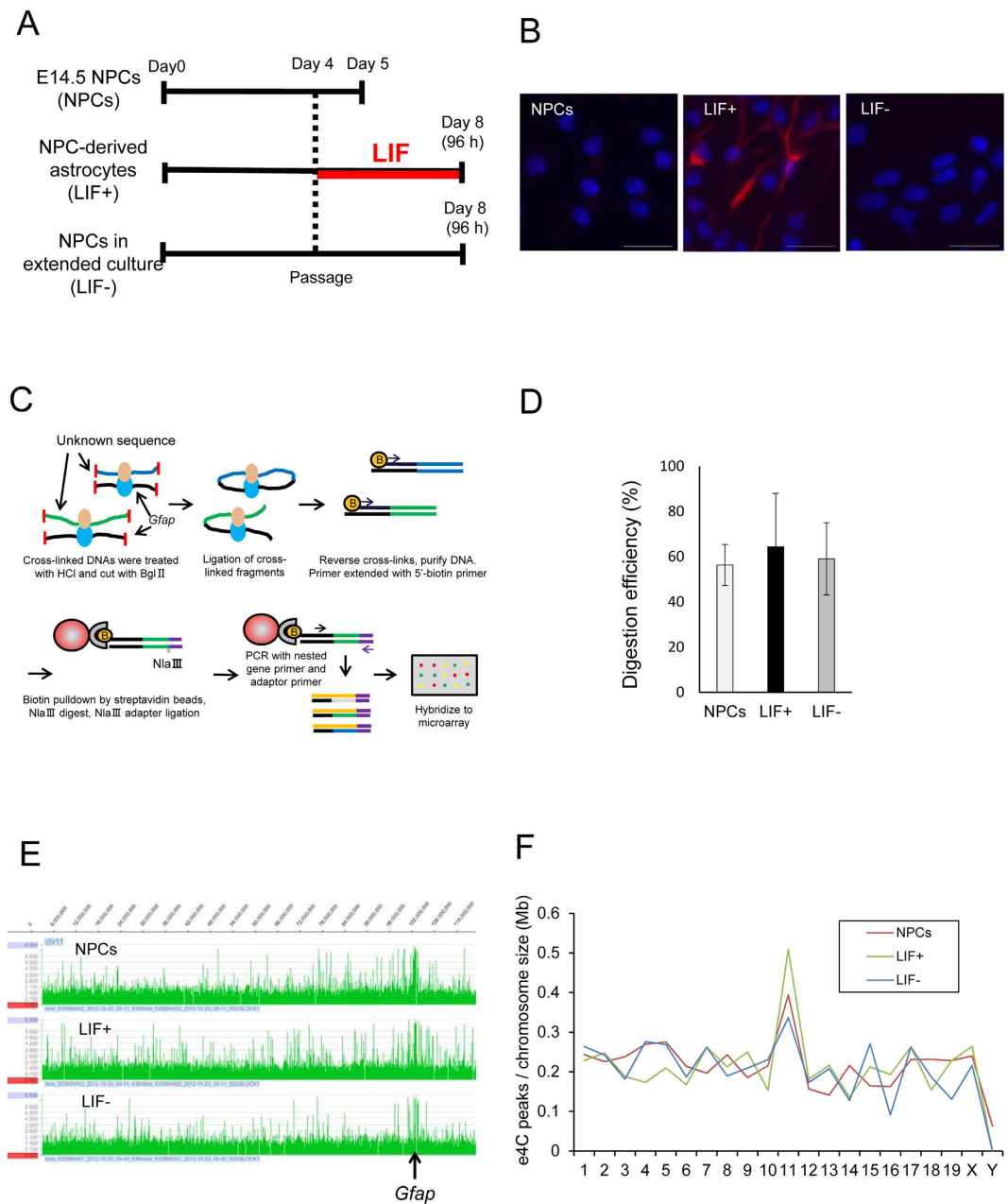


Figure 1. Genome-wide interactions of the *Gfap* loci in NPCs, LIF+, and LIF– cells. (A) Schematic experimental protocol. NPCs isolated from E14.5 mouse telencephalon were cultured and replated on day 4. On day 5, cells were used for experiments as NPCs. NPC-derived astrocytes and NPCs in extended culture were collected after an additional 4 days of culture with or without LIF. On day 8, the cells were used for experiments as LIF+ or LIF– cells. **(B)** NPCs, LIF+, and LIF– cells were stained with an anti-*GFAP* antibody (red, *GFAP*; blue, DAPI). Scale bar = 20 μ m. **(C)** Schematic representation of the modified e4C method. Chromatin was fixed in paraformaldehyde and treated with 0.1 N hydrochloric acid, then digested with *Bgl*II and ligated. The resulting hybrid molecules were used as a template for the primer extension reaction using the bait-specific primer. This was followed by adaptor ligation, nested PCR, labelling with fluorescent dye, and hybridization to a custom microarray. **(D)** The digestion efficiency of the DNA samples used for modified e4C. Relative amounts of PCR products on a *Bgl*III recognition site located near the *Gfap* STAT3 binding site (GSBS) are shown. **(E)** Association profiles were determined as the signal ratio of e4C samples to reference genomic DNA. $\text{Log}_2(\text{e4C DNA/genomic DNA}) = 2$ was set as a cut-off value. **(F)** Number of e4C peaks on each chromosome. Chromosome sizes were obtained from the Mouse Genome Browser Gateway (NCBI37/mm9).

were observed for *A2m* and *Gab1*. The number of cells with close proximity to *Gfap* did not change significantly for *Ahnak* and *Nme8* in different types of cells. Overall, there were no significant differences in the distribution of distances between *Gfap* and those gene loci among the three types of cells (Figure S4). We did notice a

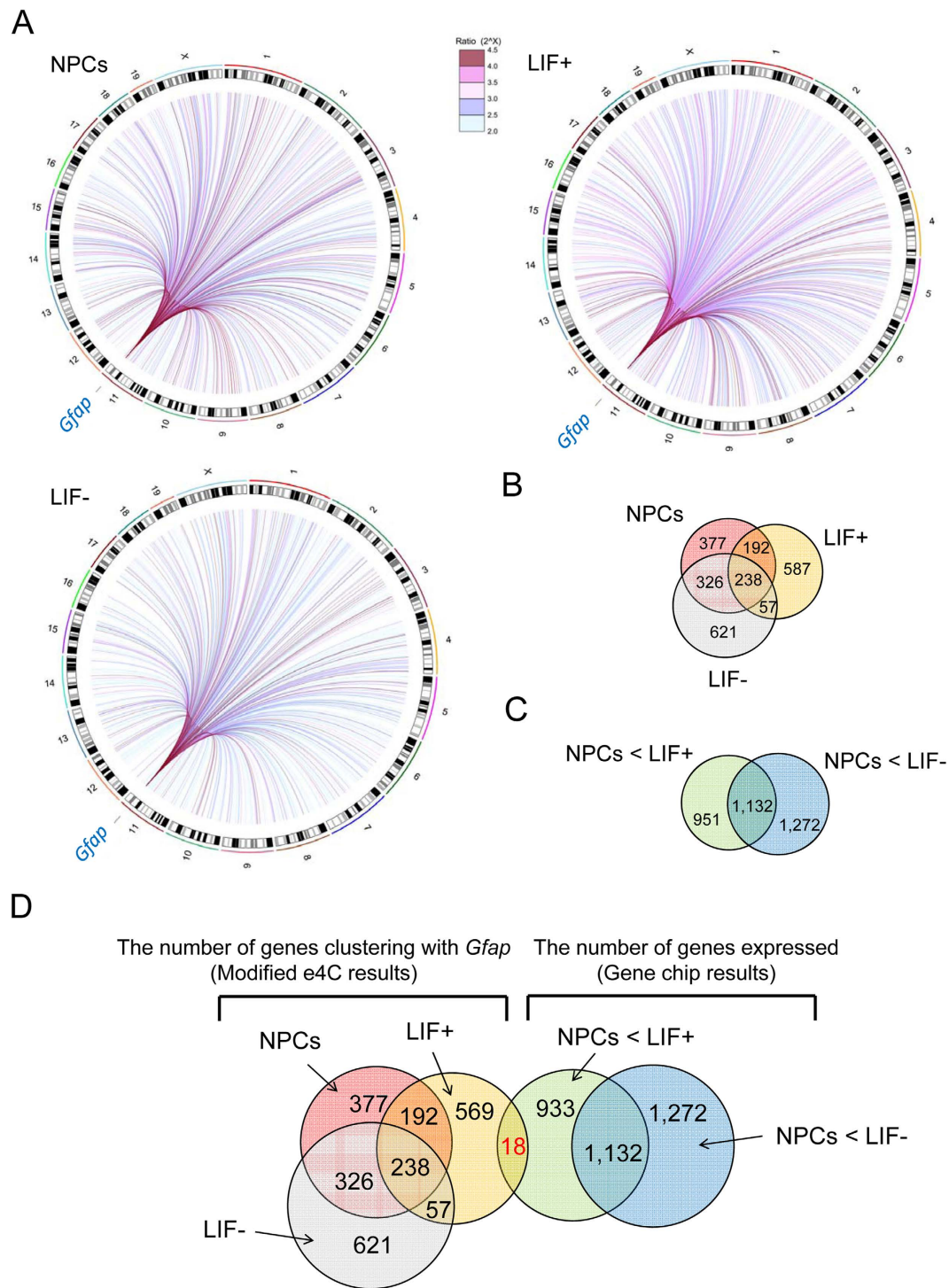


Figure 2. Identification of putative genes specifically associated with *Gfap* and expressed in LIF+ cells. (A) Circos plots showing interactions between the *Gfap* locus and interacting partners, with each line representing an interaction. Chromosomes are plotted along the circle. These plots were generated using the R package. The results of NPCs, LIF+, and LIF- cells are shown separately. Each line color represents the ratios of intensities of peaks compared to genomic controls. (B) Venn diagram of the *Gfap*-associating *Bgl*II fragments from e4C results in NPCs, LIF+, and LIF- cells. (C) Venn diagram of the genes expressed more highly in either LIF+ or LIF- cells than NPCs. (D) Venn diagram of 587 genes specifically associated with *Gfap* in LIF+ cells and 946 genes specifically expressed in LIF+ cells.

consistent difference between NPCs and LIF- in the association of each pair of probes (Fig. 4B). This was not explained by the nuclear diameter because it was not significantly different between NPCs, LIF+, and LIF- cells

Accession	Gene symbol	NPCs	LIF+	LIF–	LIF+/NPCs	LIF–/NPCs	STAT3 binding site	Distance to probe region detected by e4C
NM_011019	<i>Osmr</i>	0.20	14.43	0.27	72.14	1.33	○	10–30 kb from TES
NM_008760	<i>Ogn</i>	0.10	3.46	0.04	34.58	0.42	○	10–30 kb from TSS
NM_011313	<i>S100a6</i>	1.20	29.50	1.27	24.58	1.06	○	10–30 kb from TES
NM_172471	<i>Itih5</i>	0.60	11.04	1.81	18.40	3.02	○	>30 kb from TES
NM_175459	<i>Glis3</i>	0.40	2.97	0.38	7.43	0.94	○	>30 kb from TSS
NM_008046	<i>Fst</i>	0.20	1.36	0.23	6.82	1.16	○	>30 kb from TES
NM_133832	<i>Rdh10</i>	1.20	4.97	1.76	4.14	1.47	○	5–10 kb from TSS
NM_024283	<i>Ecrf4</i>	0.50	1.76	0.08	3.52	0.16	○	>30 kb from TSS
NM_011883	<i>Rnfl3</i>	2.70	9.12	5.38	3.38	1.99	○	10–30 kb from TES
NM_175628	<i>A2m</i>	11.10	37.32	4.31	3.36	0.39	○	5–10 kb from TES
BC019423	<i>Rsph9</i>	0.80	1.99	0.61	2.48	0.76	×	10–30 kb from TES
BC110634	<i>Galnt1</i>	3.00	7.42	3.95	2.47	1.32	○	>30 kb from TES
NM_009011	<i>Rad23b</i>	6.00	14.72	11.27	2.45	1.88	×	>30 kb from TES
NM_009371	<i>Tgfb2</i>	1.00	2.33	0.29	2.33	0.29	○	>30 kb from TSS
NM_173876	<i>Cln3</i>	3.30	7.38	6.31	2.24	1.91	○	<2 kb from TES
NM_175836	<i>Spnb2</i>	4.60	10.06	6.84	2.19	1.49	○	>30 kb from TSS
NM_008654	<i>Ppp1r15a</i>	1.00	2.14	1.34	2.14	1.34	×	>30 kb from TES
BC094659	<i>Gab1</i>	11.70	23.66	11.72	2.02	1.00	×	2–5 kb from TES

Table 1. List of the 18 putative genes specifically associated with *Gfap* and expressed in LIF+ cells. TSS; transcription start site, TES; transcription end site.

($10.4 \pm 1.5 \mu\text{m}$ in NPCs, $10.1 \pm 1.1 \mu\text{m}$ in LIF+, and $10.2 \pm 1.2 \mu\text{m}$ in LIF–). LIF– is an extended culture of NPCs, which readily differentiate into astrocytes with changes in epigenetic programs dedicated to astrocytes^{19,33}. This might explain the constant difference between NPCs and LIF–. Consequently, these cell-based analysis results agree with the e4C findings.

Detailed gene association analysis. We previously reported that *Gfap* is expressed in a random monoallelic manner in LIF+ cells and cortical astrocytes⁶. Hence, we investigated whether the alleles that associate are preferentially expressed with a simultaneous RNA/DNA-FISH assay. We used a probe targeting mature *Gfap* transcripts and probes targeting *Gfap* and *Ogn* gene loci. A significantly larger number of active *Gfap* alleles associated with *Ogn* loci than inactive ones (Fig. 5A,B). The results suggest that *Gfap* transcriptional activity is at least partially correlated with gene association.

According to the position weight matrix of STAT3 derived from the previous ChIP-seq data³⁴, there are STAT3 binding motifs within 5 kb of the transcription start sites of 14 of the 18 genes screened with the e4C (Table 1). This indicates that these genes are activated by the common transcription factor, STAT3. We therefore aimed to determine whether the genes with STAT3 binding motifs share the same locale, i.e. a transcription factory. We performed a multicolor DNA FISH assay using probes targeting *Gfap*, *Ogn*, and *Osmr*. We did not observe these three genes clustered in our system (Fig. 5C,D), indicating that these gene associations did not simultaneously occur.

In addition, a detailed analysis of the e4C results did not indicate direct interaction among the promoters of those genes; positive probes for those 18 genes were not mapped on the promoter regions (Table 1).

Discussion

In this study, we identified 18 genes that are specifically expressed and associated with the astrocyte-specific gene *Gfap* in NPC-derived astrocytes (LIF+ cells) by comparing gene association data obtained by applying a modified e4C genome-wide screening method to genome-wide gene expression profiles obtained from a microarray analysis. Importantly, these results were reproducible with two additional independent and distinct methods: DNA FISH and RT-qPCR. We also showed that the association of these genes correlated with *Gfap* transcriptional activity (Fig. 5A,B).

Long-range chromatin interactions including gene clustering are increasingly being recognized as an important mechanism to regulate gene expression. In many instances, these multi-gene complexes are hypothesized to be organized within “transcription factories” containing RNA polymerase II (RNAPII) and other components involved in transcription^{5,8,35}. Furthermore, most genes in these transcription factories are transcribed cooperatively, and some of these genes can influence each other³⁶. Although we need to investigate further to reveal the functional significance of clustering, the 18 genes identified in this study may have roles in regulating *Gfap* expression and astrocyte differentiation. In fact, one of the identified genes, *Osmr*, encodes an oncostatin M (OSM) receptor known to be involved in STAT3 activation and subsequent *Gfap* expression during astrocyte differentiation³⁷. The OSMR is essential for OSM, a member of the interleukin (IL)-6 family of cytokines, to activate downstream JAK-STAT signaling pathways by forming a heterodimer with the common signal transducer gp130. Interestingly, *Osmr* itself is transcriptionally activated by STAT3^{38,39}. This suggests that STAT3 may be involved in gene associations between *Gfap* and *Osmr*, and that the association has roles in initiating enhanceosomes for

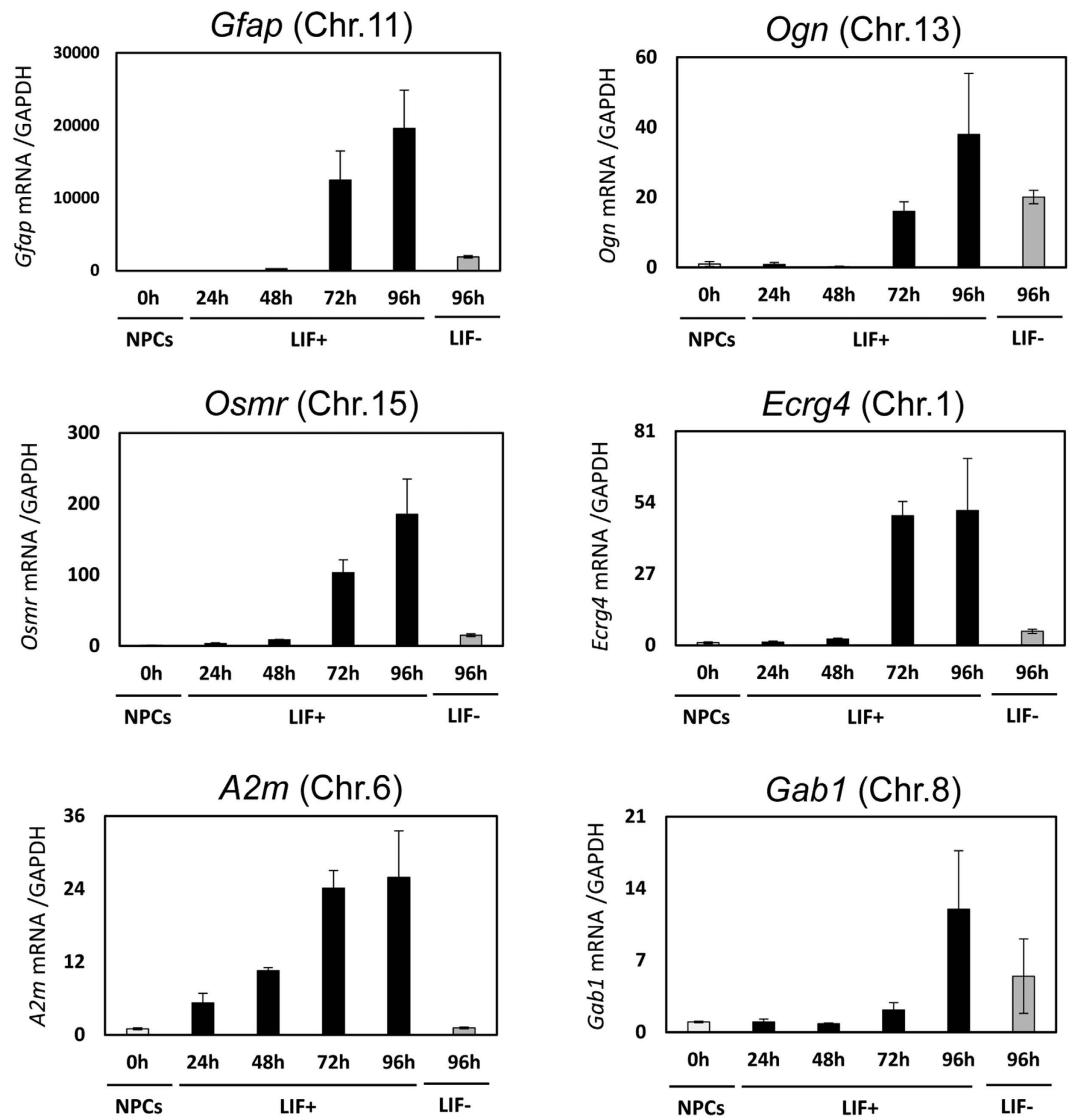
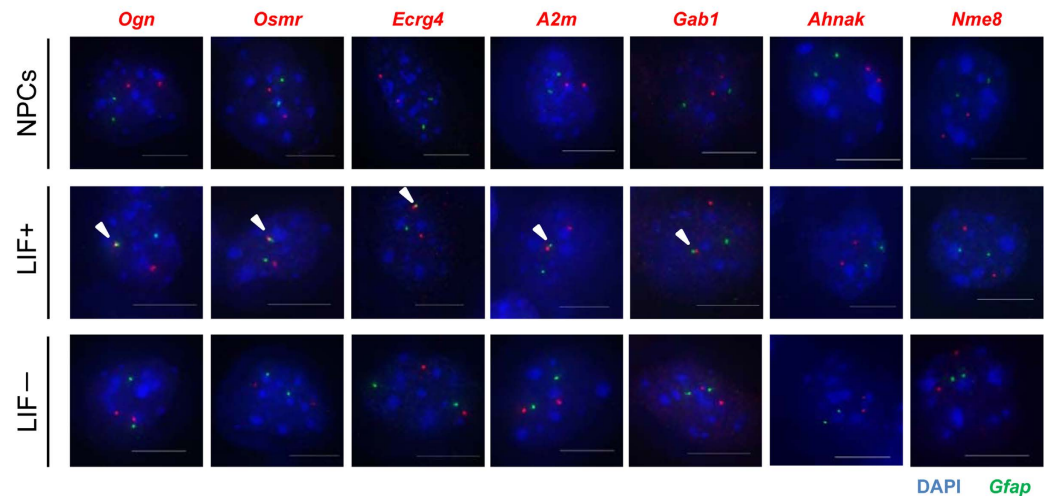


Figure 3. Validation in expression of *Gfap* and five putative genes identified by Genechip and modified e4C. Quantitative RT-PCR was performed on *Gfap* and five genes selected from modified e4C. The expression level of each gene was determined in NPCs (NPCs 0 h), cells stimulated with LIF for different periods of time (LIF+ 24 h, 48 h, 72 h, 96 h), and without LIF (LIF- 96 h). The results were normalized to *Gapdh* expression. Each graph represents the mean (\pm SEM) relative amount (compared with NPCs) in at least three experiments.

Gfap expression and astrocyte differentiation. Another identified gene, *Ecrg4*, participates in NPC cell cycle arrest through a mechanism involving the proteasome-dependent degradation of cyclins D1 and D3^{27,40}. Gene associations between *Gfap* and *Ecrg4* may be relevant for the appropriate timing of transcription initiation and astrocyte differentiation. On the other hand, e4C-positive probes were not found around the STAT3 binding sites of the 14 genes that possessed them. This indicates that close proximity of those genes, but not direct association of STAT3 binding sites, may play a role in enhancing transcription. It might be interesting to perform higher molecular resolution tests such as sequential chromatin immunoprecipitation to study gene interactions.

Changes in gene clustering reflect a wide range of genome functions such as replication, imprinting, and transcription⁴¹⁻⁴³. Although we revealed transcription-related gene associations of *Gfap*, transcription is not the only factor governing associations between *Gfap* and other genes. *Gfap* is monoallelically expressed in primary astrocytes and asymmetrically replicated, as are a number of other monoallelically expressed genes⁶. Furthermore, the reduction of transcription-repressive histone modifications (e.g., methylation of H3 at lysine 9 [H3K9me2, 3]) and the expansion of transcription-permissive histone modifications (e.g., methylation of H3 at lysine 4 [H3K4me3] and acetylation of H3 at lysine 9 [H3K9Ac] at GSBS) are important for *Gfap* expression⁴⁴⁻⁴⁶. These previous findings support the idea that changes in the pairing of associated genes may depend not only on transcriptional activities but also on replication timing and epigenetic modifications. In fact, during differentiation of embryonic stem cells (ESCs) to NPCs, switching of chromosomal domains between early and late replication in the S-phase results in changes in gene pairings⁴³. Furthermore, proteins that bind to H3K27me3 and cause DNA

A



B

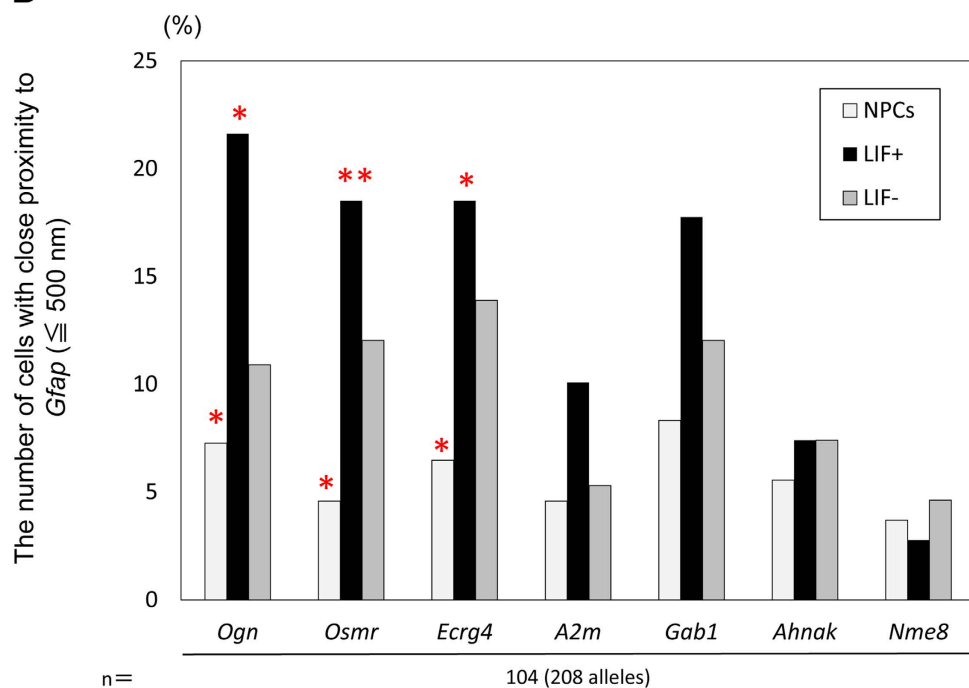


Figure 4. Validation of gene association between *Gfap* and five putative genes identified by Genechip and modified e4C. (A) Projected images of double-labeled DNA FISH in NPCs, LIF+, and LIF- cells for *Gfap* (green) and other genes (red). Nuclei were counterstained with DAPI (blue). Scale bar = 5 μ m. **(B)** Association frequencies determined with DNA FISH for the indicated gene pairs in NPCs, LIF+, and LIF- cells. ** $P < 0.01$, * $P < 0.05$. **(C)** Nuclear size as measured by DAPI staining in NPCs, LIF+, and LIF- cells. ** $P < 0.01$, * $P < 0.05$.

methylation can enhance gene associations and change transcription states⁴⁷. Thus, it would be interesting to investigate replication timing and genome-wide epigenetic modifications in NPCs and NPC-derived astrocytes. In this sense, it will also be interesting to identify genes that associate with *Gfap* in NPCs from mouse telencephalon at mid-gestation (e.g., E11.5) because the promoter region of *Gfap* is highly DNA methylated and H3K27 is tri-methylated to maintain a transcriptionally repressed state^{19,48}.

Recent studies that couple 3C derivatives and deep sequencing have shown that the genome's spatial organization is more complex than initially thought. Dixon *et al.*⁴⁹ showed that the genome is organized into large, discrete, and self-interacting domains and termed these "topological domains." These serve as a fundamental

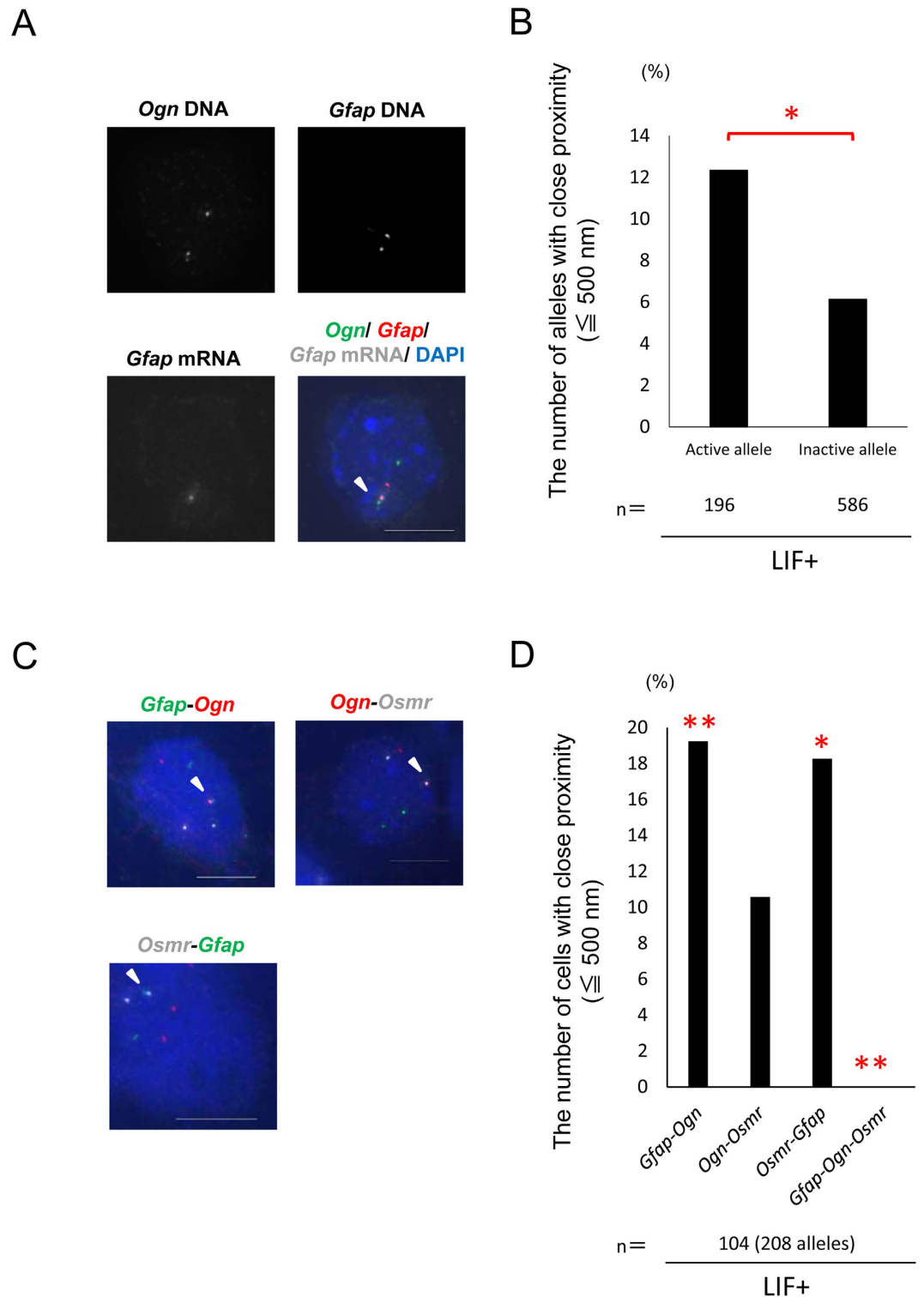


Figure 5. Detailed gene association analysis. (A) Projected images of triple-labeled RNA/DNA FISH in LIF+ cells for *Ogn* DNA (green), *Gfap* RNA (red), and *Gfap* DNA (white). Nuclei were counterstained with DAPI (blue). Scale bar = 5 μ m. (B) Association frequencies of *Gfap* active alleles and inactive alleles determined with RNA/DNA FISH for LIF+ cells. * $P < 0.05$. (C) Projected images of triple-labeled DNA FISH in LIF+ cells for *Gfap* (green), *Ogn* (red), and *Osmr* (white). Nuclei were counterstained with DAPI (blue). Scale bar = 5 μ m. (D) Association frequencies determined with DNA FISH for the indicated gene pairs in LIF+ cells. ** $P < 0.01$, * $P < 0.05$.

organizational framework of the genome because the broader organization of these topological domains is largely unchanged during differentiation, and structural changes mostly occur within domains^{49,50}. In this study, we found both specifically associating genes and genes that stably associate with *Gfap*. Although their molecular functions and significances remain unknown, they presumably act as the boundaries of topological domain-like structures and may play a role in cell-type-specific gene associations and expression. Among several such factors that participate in the organization of a higher-order chromatin structure is the CCCTC-binding factor (CTCF), which is enriched at the boundaries of topological domains⁴⁹. In addition, the loss of cohesin, which co-localizes with CTCF, leads to global perturbation of topological domain organization and transcriptional regulation in NPCs and NPC-derived astrocytes^{51,52}. Cohesin is essential for gene expression in neural cells, and its dysfunction leads to Cornelia de Lange Syndrome (CdLS), which presents as congenital anomalies and mental retardation^{53,54}. It would be interesting to map CTCF onto the genome by using chromatin immunoprecipitation in our culture system.

Several concerns regarding the 3C and its derivative techniques have been pointed out in recent publications^{24,32,55}. One is that 3C ligation products largely originate from insoluble rather than soluble fractions of chromatin⁵⁶. The results are therefore influenced by nuclear compartment or chromatin folding. Another issue is that the ability of sequences to become cross-linked and captured to distant sequences by Hi-C corresponds to the looping-out frequency from chromosome territories⁵⁷. This indicates that the results could be affected by differences in digestion efficiency with restriction endonucleases. The modified-e4C assay with hydrochloric acid treatment used in this paper likely ameliorated this problem. Nevertheless, 3C and its derivative techniques need to be validated by completely independent methods such as FISH because the results do not always simply represent spatial proximity or molecular interaction. In addition, these methods assess millions of cells and estimate an average chromatin conformation, which prompted us to ask whether multiple identified genes simultaneously share the same nuclear locale. Three-color DNA FISH (Fig. 5C,D) showed that at least the selected three genes were rarely present in the same locale of the nucleus simultaneously. Conclusively, our results suggest that verification of results obtained with 3C and its derivatives by FISH is indispensable to give accurate insights into the nature of gene clustering. In summary, we identified genes that specifically associate with the *Gfap* gene locus and are expressed in NPC-derived astrocytes. These results suggest that transcription of one of the astrocyte-specific genes, *Gfap*, is cooperatively regulated by co-expressed genes and their regulatory factors. We provide a new framework for *Gfap* expression and astrocyte differentiation that will help uncover the precise mechanisms of *Gfap* expression and astrocyte differentiation.

Methods

Cell culture. Pregnant ICR mice were used to prepare NPCs. The experimental protocols described below were performed according to the animal experimentation guidelines of Gunma University. NPCs prepared from the telencephalon of ICR mice at embryonic day (E) 14.5 were cultured as previously described¹⁹. Briefly, the telencephalon was triturated in Hanks' balanced salt solution by gently pipetting with 1-mL pipette tips. Dissociated cells were cultured in N2-supplemented Dulbecco's Modified Eagle's Medium with F12 containing 10 ng/mL basic fibroblast growth factor (bFGF; R&D Systems, Minneapolis, MN) on culture dishes (Corning, Corning, NY) that were pre-coated with poly-L-ornithine and fibronectin (Sigma-Aldrich, St. Louis, MO). For astrocyte differentiation, the cells were re-plated on fibronectin/poly-L-Lysine-coated glass coverslips (MATSUNAMI, Osaka, Japan) or culture dishes that were pre-coated with poly-L-ornithine and fibronectin after 4 d of culture and were stimulated for 4 d in the presence of LIF (50 ng/mL; Millipore, Billerica, MA). All animal procedures were conducted with the approval of Gunma University Animal Care and Use Committee and were in full compliance with the Committee's guidelines.

Immunocytochemistry. Cells cultured on coated glass coverslips were fixed in 4% paraformaldehyde in phosphate-buffered saline (PBS) and washed with PBS as described previously⁶. A mouse monoclonal antibody against GFAP (Sigma, #G-6171) was used as a primary antibody. Alexa 555- or Alexa 647-conjugated secondary antibodies (Invitrogen, Carlsbad, CA) were used for visualization. For simultaneous labeling experiments, immunostaining was performed after FISH.

Modified e4C. e4C was performed as described previously⁸ with minor modifications. Briefly, cells were exposed to 2% formaldehyde for 10 min. Cells were collected after they were quenched with 125 mM glycine. After they were homogenized, the cells were lysed in 2 mL lysis buffer (10 mM Tris-HCl [pH 7.5], 10 mM NaCl, 0.2% NP-40, 1× protease inhibitor cocktail [Nacalai Tesque]) for 1.5 h at 4 °C and centrifuged to remove the supernatant. Extracted nuclei were treated with 0.1 N HCl for 10 min and neutralized with 0.1 N NaOH and then incubated with 850 U *Bgl*III (New England Biolabs, Ipswich, MA) overnight at 37 °C. After being inactivated in 1.6% SDS at 65 °C for 20 min, samples were diluted in 4.8 mL ligation buffer (66 mM Tris-HCl [pH 7.5], 6.6 mM MgCl₂, 10 mM DTT, 0.1 mM ATP) and 2400 U T4 ligase (New England Biolabs) and incubated at 16 °C for 4 h. Ligated chromatin was digested by proteinase K (100 ng/mL; Merck, White House Station, NJ). After phenol-chloroform extraction, DNA was ethanol precipitated. Then, digestion efficiency was verified as previously described⁵⁸, and the 3C products were processed for primer extension with 2 U Vent (exo-) DNA polymerase (New England Biolabs) and biotinylated bait-region-specific primers. After being bound to streptavidin-coated magnetic beads (Dynabeads M-280, Invitrogen), the biotinylated products on beads were digested with 20 U *Nla*III (New England Biolabs), followed by an adaptor ligation with 2000 U T4 ligase (New England Biolabs). The beads were used for PCR with nested bait-region-specific primers and adaptor-specific primers. The amplified products were digested again with 20 U *Bgl*II and 20 U *Nla*III (New England Biolabs). Following phenol-chloroform extraction, the e4C products were ethanol precipitated. The samples were hybridized to the customized microarray following NimbleGen's protocol. The following primers

and adapters were used: 5' GGACATGATGAGGTCCAGTC 3' and 5' GCTTGCTGAGGTTCTCCTAATG 3', 5' GCCCAGAGTACTCACCTTG 3' and 5' CCAGGATGCCAGGATGTCAG 3' (*Gfap* *Bgl*III site and GSBS for Digestion efficiency check), 5' biotin GCTTGCTGAGGTTCTCCTAATG 3' (biotinylated bait-specific primer for primer extension), 5' TTGGATTTGCTGGTGCAGTACAACACTAGGCTTAATAGGGACATG 3' and 5' phosphorylated CCCTATTAAGCCTAGTTGTACTGCACCAGCAAATCC 3' amine C7 (*Nla* adapter strands), and 5' GGATTTGCTGGTGCAGTACA 3' and 5' GAATAATGGCATAAGTGAGGGAG 3' (for nested PCR).

e4C microarray. e4C material was labeled and competitively hybridized with digested mouse genomic DNA as described previously⁸. The customized NimbleGen array consists of 1.4 million probes with a length of 50–70mer covering as many fragments as possible with *Bgl*III and *Nla*III sites on the 5' and 3' ends, respectively, and a size >100bp based on NCBI37/mm9. Two biological replicates were performed for each condition, and reproducible positive probes were identified as e4C hits by setting a cut-off value of \log_2 (e4C signal/genomic control) = 2 (200-bp sliding window). e4C genes were identified by mapping all genes within 50 kb from each peak. Circos Plots of the results were generated with R version 3.3.3 using Package RCircos ver1.1.2.⁵⁹

Sample preparation and GeneChip analysis. Sample preparation and GeneChip analysis were performed as described previously⁶⁰. The expression data were converted to copy numbers of mRNA per cell by the Percellome method, quality controlled, and analyzed using Percellome software²⁵. Genes with copy numbers that increased by at least two-fold were identified as upregulated genes, while genes whose copy numbers of mRNA were <1 were excluded.

FISH. Probes for DNA FISH were generated by nick translation of BAC clones covering genes of interest (BACPAC Resources). The following BAC clones were used: RP24–155G1 (*Gfap*), RP24–152H11 (*Ogn*), RP23–198P20 (*Osmr*), RP23–185E14 (*A2m*), RP24–214M12 (*Ecr4*), RP24–114L21 (*Gab1*), RP23–118O2 (*Ahnak*), and RP23–211E13 (*Nme8*). FISH was essentially performed as described previously⁶. Briefly, cells were fixed with 4% PFA and kept in PBS at 4°C until use. The cells were permeabilized with 0.5% Triton X-100/PBS and treated with 0.1 N HCl for 10 min. Cells were denatured for 30 min at room temperature in 50% formamide with 2× SSC. Hybridization was performed overnight at 37°C with dinitrophenyl (DNP) or digoxigenin (DIG) or biotin-labeled probes and detected with Alexa 488-conjugated anti-DNP (Invitrogen) or rhodamine-conjugated anti-DIG antibody (Roche, Basel, Switzerland) or Alexa 647-conjugated streptavidin (Invitrogen). For RNA/DNA FISH, RNA probes were made as previously described⁶. In brief, *Gfap* exon sequences were amplified using cDNA prepared from astrocytes as a template. Amplified DNA was *in vitro* transcribed and then reverse-transcribed in the presence of biotin-labeled dUTP (Invitrogen). The single-stranded biotin-labeled cDNA probe was used to detect *Gfap* mRNA. Cells were fixed with 4% PFA containing 10% acetic acid and kept in 70% ethanol at –30°C until use. Cells were digested with 0.05% pepsin/0.01 N HCl, dehydrated through ethanol treatment, and hybridized to the single-stranded DNA probe against *Gfap* cDNA. RNA-probe hybrids were detected with horseradish peroxidase-conjugated streptavidin and further labeled with Alexa598-conjugated tyramide using the TSA kit (Invitrogen). DNA FISH was performed after RNase treatment.

Microscopy and image analysis. A DeltaVision microscope (CORNES Technologies) was used to analyze the results of DNA FISH. 3D images were obtained from serial Z-sections of 8.0-μm thickness in 0.1-μm intervals. For association analysis, the shortest distances between two FISH signals were examined by softWoRx Explorer1.3 (Applied Precision). Genes were considered associated when they were positioned within 500 nm of *Gfap*.

Quantitative RT-PCR. Total RNAs were extracted with ToTALLY RNA™ Total RNA Isolation Kit and then treated with DNaseI (Life Technologies). Complementary DNAs were synthesized from 2 μg total RNA using Superscript II (Life Technologies). Quantitative real-time PCR (qPCR) was performed in an Applied Biosystems 7900HT Fast Real-Time PCR system (Life Technologies) using the KAPA SYBR Fast qPCR Kit (Kapa Biosystems, Woburn, MA). Expression of the target genes was normalized to that of glyceraldehyde-3-phosphate dehydrogenase (*Gapdh*). The following primers were used: 5' ATCGAGATCGCCACCTACAG 3' and 5' CTCACATC ACCACGTCCTTG 3' (for *Gfap*), 5' TGCAACAGGCAATTCTGAAG 3', 5' TGCAACAGGCAATTCTGAAG 3' and 5' TCCTTGCGAGTCAGCTTTTT 3' (for *Ogn*), 5' ACACCAAGTCCCTTCCACAG 3' and 5' ATGGTGACATTGG AGCCTTC 3' (for *Osmr*), 5' GCCTGAGGTACAGCAGTGGT 3' and 5' ATGGCCGCATCTTCATCATA 3' (for *Ecr4*), 5' CTTCTATTATCTGATGATGGCAAAGG 3' and 5' CCTGCGTCACAGGCAGAAC 3' (for *A2m*), 5' CCAGGACG ATCCACAAGACT 3' and 5' TTCATTCCGTGTTTGCTCTG 3' (for *Gab1*), 5' ACCACAGTCCATGCCATCAC 3' and 5' TCCACCACCCTGTTGCTGTA 3' (for *Gapdh*).

Statistical analyses. Residual analyses of chi-squared tests were used for Figs 4B and 5D. Fisher's and Kolmogorov-Smirnov tests were used for Fig. 5B and Figure S4, respectively. One-way analysis of variance (ANOVA) with nonparametric tests was used to compare nuclei sizes (data not shown).

Accession codes. Data are deposited in NCBI's Gene Expression Omnibus and are accessible through GEO Series accession number GSE66097.

References

- Misteli, T. The cell biology of genomes: bringing the double helix to life. *Cell* **152**, 1209–1212 (2013).
- Cavalli, G. & Misteli, T. Functional implications of genome topology. *Nat Struct Mol Biol* **20**, 290–299 (2013).
- Bantignies, F. *et al.* Polycomb-dependent regulatory contacts between distant Hox loci in *Drosophila*. *Cell* **144**, 214–226 (2011).

4. Kohwi, M., Lupton, J. R., Lai, S. L., Miller, M. R. & Doe, C. Q. Developmentally regulated subnuclear genome reorganization restricts neural progenitor competence in *Drosophila*. *Cell* **152**, 97–108 (2013).
5. Osborne, C. S. *et al.* Active genes dynamically colocalize to shared sites of ongoing transcription. *Nat Genet* **36**, 1065–1071 (2004).
6. Takizawa, T., Gudla, P. R., Guo, L., Lockett, S. & Misteli, T. Allele-specific nuclear positioning of the monoallelically expressed astrocyte marker *GFAP*. *Genes Dev* **22**, 489–498 (2008).
7. Spilianakis, C. G., Lalioti, M. D., Town, T., Lee, G. R. & Flavell, R. A. Interchromosomal associations between alternatively expressed loci. *Nature* **435**, 637–645 (2005).
8. Schoenfelder, S. *et al.* Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. *Nat Genet* **42**, 53–61 (2010).
9. Apostolou, E. *et al.* Genome-wide chromatin interactions of the Nanog locus in pluripotency, differentiation, and reprogramming. *Cell Stem Cell* **12**, 699–712 (2013).
10. Handoko, L. *et al.* CTCF-mediated functional chromatin interactome in pluripotent cells. *Nat Genet* **43**, 630–638 (2011).
11. Wang, K. C. *et al.* A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* **472**, 120–124 (2011).
12. Watanabe, T. *et al.* Higher-order chromatin regulation and differential gene expression in the human tumor necrosis factor/lymphotoxin locus in hepatocellular carcinoma cells. *Mol Cell Biol* **32**, 1529–1541 (2012).
13. de Wit, E. & de Laat, W. A decade of 3C technologies: insights into nuclear organization. *Genes Dev* **26**, 11–24 (2012).
14. Li, M., Liu, G. H. & Izpisua Belmonte, J. C. Navigating the epigenetic landscape of pluripotent stem cells. *Nat Rev Mol Cell Biol* **13**, 524–535 (2012).
15. Temple, S. The development of neural stem cells. *Nature* **414**, 112–117, doi: 10.1038/35102174 (2001).
16. Edlund, T. & Jessell, T. M. Progression from extrinsic to intrinsic signaling in cell fate specification: a view from the nervous system. *Cell* **96**, 211–224 (1999).
17. Hsieh, J. & Gage, F. H. Epigenetic control of neural stem cell fate. *Curr Opin Genet Dev* **14**, 461–469 (2004).
18. Hatada, I. *et al.* Astrocyte-specific genes are generally demethylated in neural precursor cells prior to astrocytic differentiation. *PLoS One* **3**, e3189 (2008).
19. Takizawa, T. *et al.* DNA methylation is a critical cell-intrinsic determinant of astrocyte differentiation in the fetal brain. *Dev Cell* **1**, 749–758 (2001).
20. Nakashima, K. *et al.* Synergistic signaling in fetal brain by STAT3-Smad1 complex bridged by p300. *Science* **284**, 479–482 (1999).
21. Nakashima, K. *et al.* Developmental requirement of gp130 signaling in neuronal survival and astrocyte differentiation. *J Neurosci* **19**, 5429–5434 (1999).
22. Urayama, S. *et al.* Chromatin accessibility at a STAT3 target site is altered prior to astrocyte differentiation. *Cell Struct Funct* **38**, 55–66 (2013).
23. Peric-Hupkes, D. *et al.* Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. *Mol Cell* **38**, 603–613 (2010).
24. Belmont, A. S. Large-scale chromatin organization: the good, the surprising, and the still perplexing. *Curr Opin Cell Biol* **26**, 69–78 (2014).
25. Kanno, J. *et al.* “Per cell” normalization method for mRNA measurement by quantitative PCR and microarrays. *BMC Genomics* **7**, 64 (2006).
26. Duchemin, A. M., Ren, Q., Neff, N. H. & Hadjiconstantinou, M. GM1-induced activation of phosphatidylinositol 3-kinase: involvement of Trk receptors. *J Neurochem* **104**, 1466–1477 (2008).
27. Gonzalez, A. M. *et al.* *Ecr4* expression and its product augurin in the choroid plexus: impact on fetal brain development, cerebrospinal fluid homeostasis and neuroprogenitor cell response to CNS injury. *Fluids Barriers CNS* **8**, 6 (2011).
28. Jeong, E. Y. *et al.* Enhancement of IGF-2-induced neurite outgrowth by IGF-binding protein-2 and osteoglycin in SH-SY5Y human neuroblastoma cells. *Neurosci Lett* **548**, 249–254 (2013).
29. Kovacs, D. M. Alpha2-macroglobulin in late-onset Alzheimer’s disease. *Exp Gerontol* **35**, 473–479 (2000).
30. Van Wagoner, N. J., Choi, C., Repovic, P. & Benveniste, E. N. Oncostatin M regulation of interleukin-6 expression in astrocytes: biphasic regulation involving the mitogen-activated protein kinases ERK1/2 and p38. *J Neurochem* **75**, 563–575 (2000).
31. Simonis, M., Kooren, J. & de Laat, W. An evaluation of 3C-based methods to capture DNA interactions. *Nat Methods* **4**, 895–901 (2007).
32. Williamson, I. *et al.* Spatial genome organization: contrasting views from chromosome conformation capture and fluorescence *in situ* hybridization. *Genes Dev* **28**, 2778–2791 (2014).
33. 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* **572**, 184–188 (2004).
34. Chen, X. *et al.* Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* **133**, 1106–1117 (2008).
35. Cook, P. R. The organization of replication and transcription. *Science* **284**, 1790–1795 (1999).
36. Li, G. *et al.* Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell* **148**, 84–98 (2012).
37. Yanagisawa, M., Nakashima, K. & Taga, T. STAT3-mediated astrocyte differentiation from mouse fetal neuroepithelial cells by mouse oncostatin M. *Neurosci Lett* **269**, 169–172 (1999).
38. Clarkson, R. W. *et al.* The genes induced by signal transducer and activators of transcription (STAT)3 and STAT5 in mammary epithelial cells define the roles of these STATs in mammary development. *Mol Endocrinol* **20**, 675–685 (2006).
39. Tiffen, P. G. *et al.* A dual role for oncostatin M signaling in the differentiation and death of mammary epithelial cells *in vivo*. *Mol Endocrinol* **22**, 2677–2688 (2008).
40. Kujuro, Y., Suzuki, N. & Kondo, T. Esophageal cancer-related gene 4 is a secreted inducer of cell senescence expressed by aged CNS precursor cells. *Proc Natl Acad Sci USA* **107**, 8259–8264 (2010).
41. Le May, N., Fradin, D., Iltis, I., Bougneres, P. & Egly, J. M. XPG and XPF endonucleases trigger chromatin looping and DNA demethylation for accurate expression of activated genes. *Mol Cell* **47**, 622–632 (2012).
42. Ling, J. Q. *et al.* CTCF mediates interchromosomal colocalization between *Igf2/H19* and *Wsb1/Nf1*. *Science* **312**, 269–272 (2006).
43. Takebayashi, S., Dileep, V., Ryba, T., Dennis, J. H. & Gilbert, D. M. Chromatin-interaction compartment switch at developmentally regulated chromosomal domains reveals an unusual principle of chromatin folding. *Proc Natl Acad Sci USA* **109**, 12574–12579 (2012).
44. Cheng, P. Y. *et al.* Interplay between SIN3A and STAT3 mediates chromatin conformational changes and *GFAP* expression during cellular differentiation. *PLoS One* **6**, e22018 (2011).
45. Song, M. R. & Ghosh, A. FGF2-induced chromatin remodeling regulates CNTF-mediated gene expression and astrocyte differentiation. *Nat Neurosci* **7**, 229–235 (2004).
46. Tan, S. L. *et al.* Essential roles of the histone methyltransferase ESET in the epigenetic control of neural progenitor cells during development. *Development* **139**, 3806–3816 (2012).
47. Tiwari, V. K. *et al.* PcG proteins, DNA methylation, and gene repression by chromatin looping. *PLoS Biol* **6**, 2911–2927 (2008).
48. Sparmann, A. *et al.* The chromodomain helicase *Chd4* is required for Polycomb-mediated inhibition of astroglial differentiation. *Embo J* **32**, 1598–1612 (2013).

49. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
50. Holwerda, S. & de Laat, W. Chromatin loops, gene positioning, and gene expression. *Front Genet* **3**, 217 (2012).
51. Wendt, K. S. *et al.* Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* **451**, 796–801 (2008).
52. Sofueva, S. *et al.* Cohesin-mediated interactions organize chromosomal domain architecture. *Embo J* **32**, 3119–3129 (2013).
53. Liu, J. & Baynam, G. Cornelia de Lange syndrome. *Adv Exp Med Biol* **685**, 111–123 (2010).
54. Chang, J. *et al.* Nicotinamide adenine dinucleotide (NAD)-regulated DNA methylation alters CCCTC-binding factor (CTCF)/cohesin binding and transcription at the BDNF locus. *Proc Natl Acad Sci USA* **107**, 21836–21841 (2010).
55. Pombo, A. & Dillon, N. Three-dimensional genome architecture: players and mechanisms. *Nat Rev Mol Cell Biol* **16**, 245–257 (2015).
56. Gavrilov, A. A. *et al.* Disclosure of a structural milieu for the proximity ligation reveals the elusive nature of an active chromatin hub. *Nucleic Acids Res* **41**, 3563–3575 (2013).
57. Kalhor, R., Tjong, H., Jayathilaka, N., Alber, F. & Chen, L. Genome architectures revealed by tethered chromosome conformation capture and population-based modeling. *Nat Biotechnol* **30**, 90–98 (2012).
58. Hagege, H. *et al.* Quantitative analysis of chromosome conformation capture assays (3C-qPCR). *Nat Protoc* **2**, 1722–1733 (2007).
59. Zhang, H., Meltzer, P. & Davis, S. RCircos: an R package for Circos 2D track plots. *BMC Bioinformatics* **14**, 244 (2013).
60. Sanosaka, T. *et al.* Identification of genes that restrict astrocyte differentiation of midgestational neural precursor cells. *Neuroscience* **155**, 780–788, 2008.06.039 (2008).

Acknowledgements

We thank all members of department of Pediatrics for their help and advice. We thank P. Fraser for kindly providing the e4C protocol. We also thank M. Nakao, N. Saito, T. Watanabe, and S. Tomita for valuable discussion and technical advice. This work was supported by a Grant-in-Aid for Scientific Research on Innovative Areas 23114714 and a grant from the JST Strategic International Research Cooperative Program, SICP.

Author Contributions

T.T. designed the experiments. K.I.¹, T.S., K.I.³ and M.O. performed the experiments. K.I.¹, T.S. and A.A. analyzed the data. K.I.¹ and T.T. wrote the manuscript. K.N., A.N., Y.U. and H.A. participated in discussions and gave valuable suggestions. All authors reviewed the paper.

Additional Information

Supplementary information accompanies this paper at <http://www.nature.com/srep>

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

How to cite this article: Ito, K. *et al.* Identification of genes associated with the astrocyte-specific gene *Gfap* during astrocyte differentiation. *Sci. Rep.* **6**, 23903; doi: 10.1038/srep23903 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>