國立台灣大學醫學院腦與心智科學研究所

碩士論文

Department of Graduate Institute of Brain and Mind Sciences

College of Medicine

National Taiwan University

Master thesis

小鼠視覺皮質層印記基因表現圖譜

Profiling of Parent-of-Origin-Specific Expression in the Mouse Visual Cortex

童雋哲

Chun-Che Tung

指導教授：黃憲松 助理教授

Advisor: Hsien-Sung Huang, Ph.D.

中華民國 103 年 7 月

July 2014

國立臺灣大學碩士學位論文
口試委員會審定書

論文中文題目

小鼠視覺皮質層印記基因表現圖譜

論文英文題目

Profiling of parent-of-origin-specific expression in the mouse visual cortex

本論文係 童雋哲 君（學號 r01454011）在國立臺灣大學腦與心智科學研究所完成之碩士學位論文，於民國 103 年 07 月 31 日承下列考試委員審查通過及口試及格，特此證明

口試委員：　　（簽名）

（指導教授）

所長　　（簽名）

中文摘要

染色體印記，是為經外遺傳的調控，而形成特殊、基因表達視其親源而定的現象。其主要發生於哺乳類的神經系統，因此，若其機制上的異常即會造成神經發育、或心智功能上的疾病。雖然染色體印記對於腦部功能是如此的重要，但印記基因明確的數目、及何為印記基因，到目前為止仍未定案。而會有如此的困難度一部份來自於腦部結構的多樣性。印記基因的表達除視其親源而定外，也會受到不同的細胞種類、發育上的時間點、以及環境的影響。為了明確的辨識印記基因，我們藉由在不同的細胞型態、以及發育時間點上，建立一篩選平台。技術上，我們利用次世代定序、以及基因轉殖鼠來達成此目標，而此篇論文所介紹，在小鼠視覺皮質層獨立興奮性神經元的方法，也可應用在不同的細胞型態上。在另一方面，為了看到環境對於印記狀態的操弄，我們選擇在完全無光線飼養下(24 小時全暗)、比較正常環境飼養(12 小時光照、12 小時黑暗)此兩種情況。藉此，破壞小鼠視覺皮層的正常發育，來看光線經驗的不同對於印記狀態的影響。藉由上述的方法，希望我的工作可以讓我們更加地了解印記基因在腦部扮演的角色、以及為何在演化上發展出如此獨立的機制。更重要的，因而印記基因異常而造成的疾病，也可藉由這樣的策略發展治療方法。

關鍵字

印記基因/視覺皮質層/外遺傳/神經系統/次世代定序

Abstract

Genomic imprinting is an epigenetic process by which certain genes are expressed in a parent-of-origin-specific manner. Genomic imprinting predominantly occurs in the mammalian nervous system. Dysfunctional genomic imprinting causes various neurological and psychiatric disorders. Despite the importance of genomic imprinting in brain function, the number and identity of imprinted genes have been proposed, but are still debated. The key obstacles to advance the field of genomic imprinting in the brain are the complex expression patterns of imprinted genes and the heterogeneous nature of the brain. The expression of imprinted genes can be regulated by different cell types, developmental stages and the environment. In order to solve the urgent problems of this field, I first developed a platform to comprehensively profile imprinted genes in a specific cell type under different developmental stages. Using the technique of deep sequencing and the advantage of engineered mice, I can determine the genomic imprinting status of excitatory neurons in mouse visual cortex. The success of the platform I developed here could apply to the profiling of the genomic imprinting status in other cell types. Regarding to the environmental effect on the imprinting status of the brain, I chose the dark rearing condition (24hr Dark) in comparison to the normal rearing (12hr/12hr, Dark/Light). Under this manipulation, we destroyed the maturation of visual cortex and determined the effect of light experience on the imprinting status of mouse visual cortex. My comprehensive profiling work of genomic imprinting status in the brain could deepen our understanding about why our brain developed genomic imprinting during evolution. More importantly, the knowledge gained here may provide therapeutic strategies for genomic imprinting disorders.

Key words

Imprinted genes/ Visual cortex/ Epigenetic/ Nervous system/ Deep sequencgin

目錄

圖目錄

Introduction

What are imprinted genes?

Despite the biallelic expression nature of most genes in mammals, approximately about 12-24% of autosomal genes had been proposed to be monoallelic expressed[1] in mice. Of such, parent-of-origin-specific expression pattern (imprinted) had been revealed in 1% of autosomal genes[2-5]. Much evidence showed that imprinted genes play major roles in embryogenesis and brain development nowadays[6-8].

To determine the parental source of the expression for imprinted genes, epigenetic markers had been set during the time of gametogenesis[9]. By marking the differentially methylated regions (DMRs), imprinted genes in the nearby locus could be controlled together by the same imprinting control center (ICR) and further form imprinting control clusters (Figure 1)[10]. Of the four well-confirmed imprinting clusters (Figure 1), black octagon represent methylated ICR, which targeted deletion in mouse have proved their roles as elements controlling parent-of-origin-specific genes expression across the whole imprinted domain. And of the same cluster, not only the protein-coding genes being imprinted but also non-coding RNAs such as miRNAs.

Specific features of imprinted genes

Other than the characteristic of clustering control, epigenetic markers also have been testified to be altered by environmental factors. Huang et al[11] affirmed the point that topoisomerase inhibitors could unsilence the paternal *Ube3a* allele. Therefore administrating the drugs (topotecan, irinotecan, etoposide and dexrazoxane) mentioned above might shed light on reactivating the functional but dormant allele of *Ube3a* in patients with Angelman syndrome[11,12].

Some genes also have their imprinting status depending on which developmental stages the organisms are at. For examples, *Commd1*[13] is one of the imprinted genes which is biallelically expressed in mouse embryonic and neonatal brains but maternally expressed (paternally imprinted) in adult mouse brain. Moreover, imprinted status could be regulated by different cell types. Taking *Ube3a* as an example, it is the gene only maternally expressed in neurons while biallelically expressed in the astrocytes[14].

The important roles of imprinted genes in the brain

Because of the important roles of imprinted genes in neurogenesis[15-17] and brain development and maturation[18-21], dysregulation, mutation, deletion or loss of active allele of imprinted genes would cause dramatic effect due to

the lack of compensating dosage. Angelmen syndrome[22,23], Prader-Willi syndrome[10,24], Autism[12,25], Ret syndrome[26,27], and Tuner syndrome[28] are the cases linking to imprinted genes. That is, people started to think about how many genes to be imprinted as well as their roles to bridge between diseases and their corresponding genes. Imprinted genes predominately occur in mammalian's nervous system, and Christopher Gregg[29,30] was the first group who screened imprinted genes on genome-wide scale in the mouse embryonic brain and adult prefrontal cortex. They mated two different strains of mouse (C57BL/6 and CAST/EiJ) to take the advantage of strain-specific single-nucleotide polymorphisms (SNPs) for identifying paternally or maternally expressed genes. After conducting RNA-seq, they found out more than 1300 protein-coding genes and putative ncRNAs associated with preferential expression of paternal or maternal allele in brains. Although they did not confirm all of the predicted imprinted genes, they created a platform for further investigation and implied more new imprinted genes which haven't been found. From then on, several studies[31,32] tried to reproduce the assays but on other brain regions or other developmental stages for the comparison. Due to the complex expression pattern of imprinted genes and heterogeneous nature of mouse brain[33], the

number and identify of imprinted genes is still debated[32].

Our hypothesis

In order to narrow down confounding factors, our lab tried to conduct single-cell-type RNA-Seq in mouse primary visual cortex. By using laser-capture microdissection (LCM) system, I am capable of cutting three different major types of cells in nervous system which had been fluorescently labeled. After cutting, I collected the cells and extracted RNA before sending out for sequencing. Using the similar strategy as Dr. Gregg did previously[29], I am able to do genome-wide screening of imprinted genes but on cellular-resolution. I can not only sequence protein-coding genes but also non-coding RNAs and miRNAs. After sequencing, I mapped back the transcript reads and determined whether they formed the unknown clusters.

Meanwhile, inspired by the evidence from Huang's previous work, I planned to identify other environmental effect on imprinting status. I compared the difference of genomic imprinting status between the dark rearing (24 hr light off) condition and standard condition with 12-h on/12-h off light cycle. Of the third relay center in visual conduction pathway, primary visual cortex had been revealed to be changed under dark rearing

condition from molecular level[34] to neuronal circuits[35]. Although the impact of light experience on genomic imprinting status is totally unknown, I am eager to determine whether these powerful manipulations could influence genomic imprinting status in the mouse visual cortex.

Material and methods

Mice

Camk2a-iCre mice (C57BL/6J background) were generated by the laboratory of G. Schütz[6] and provided by the laboratory of C.-K. J. Shen. The full strain name of Ai14 Cre reporter mice is B6.Cg-*Gt(ROSA)26Sor*$^{tm14(CAG-tdTomato)Hze}$/J (stock number: 007914) and provided by the laboratory of S.L. Lin. C57BL/6J (B6) mice (stock number: 000664) and CAST/EiJ mice (stock number: 000928) were from Jackson Laboratories. To get fluorescent excitatory neurons, we bred excitatory neuron-specific Cre-expression mice and Cre-reporter mice with a C57BL/6J (B6) background. Next, we crossed this mutant mouse (B6) with a wild-type mouse in CAST/EiJ (CAST) background [F1 initial cross (F1i); CAST wild-type mother x B6 mutant father]. This breeding produced offspring with a B6 and CAST background and red fluorescent signals in excitatory neurons. The B6/CAST mixed background has the advantage of yielding more identifiable single nucleotide polymorphisms (SNPs). At postnatal day 28 (P28), the offspring's brains were collected for further processing. To confirm our observations from F1i, we performed a reciprocal cross [F1 reciprocal cross (F1r): B6 mother x CAST father].

Genotyping PCR was performed using *Camk2a-iCre*/F (5'-CTC TGA CAG ATG CCA GGA CA-3'), *Camk2a-iCre*/R (5'-TGA TTT CAG GGA TGG ACA CA-3'), *Ai14 WT*/F (5'-AAG GGA GCT GCA GTG GAG TA-3'), *Ai14 WT*/R (5'-CCG AAA ATC TGT GGG AAG TC-3'), *Ai14 mutant*/F (5'-GGC ATT AAA GCA GCG TAT CC-3'), *Ai14 mutant*/R (5'-CTG TTC CTG TAC GGC ATG G-3'). The PCR cycling conditions were as follows: for *Camk2a-iCre* primers; initial denaturation at 94°C for 3 min followed by 35 cycles of 94°C for 30 s, 56°C for 30 s and 70°C for 60 s and a final extension at 72°C for 2 min; for *Ai14 WT and mutant* primers; initial denaturation at 94°C for 3 min followed by 35 cycles of 94°C for 20 s, 61°C for 30 s and 72°C for 30 s and a final extension at 72°C for 2 min. Amplification was performed on a C1000 Touch Thermal Cycler (Bio-Rad). Light-reared mice were raised on a 12-h light/dark cycle, whereas dark-reared mice were raised in complete darkness from birth to P28. All animal experiments were approved by the National Taiwan University College of Medicine and College of Public Health Institutional Animal Care and Use Committee (IACUC).

Immunofluorescence staining

Mice were sacrificed at P28, and brains were immersion-fixed with 4%

paraformaldehyde in 0.1 M phosphate buffer (PB) (pH 7.4) for 12 h and then cryoprotected with 30% sucrose in 0.1 M phosphate buffer at 4 °C overnight. Fixed brains were then embedded in O.C.T. compound (Surgipath, FSC 22) and sectioned into 7 μm with microtome (Leica, CM3050 S). Sections were collected and permeabilized with 0.3% Triton X-100 in 1X phosphate buffer saline (PBS) for 30 min at room temperature. Sections were further incubated with anti-CAMK2A (1:500, Acris Antibodies, AM 12015PU-S) at room temperature for 1 h, and then incubated with Alexa Fluor® 488 goat anti-mouse IgG1 (γ1) antibody (1:500, Invitrogen, A21120) and DAPI (1:10,000, Invitrogen, D-1306) for 2 h at room temperature. Sections were washed with 1XPBS containing 0.1% Triton X-100 three times (5 min/time) after primary and secondary antibodies incubation. After coverslipped with fluoromount aqueous mounting medium (Sigma, F4680), images were acquired using Zeiss LSM 780 confocal microscope.

Fluorescence-based laser capture microdissection

P28 mouse brain was cut in half coronally, immersed in RNAlater stabilization solution (Ambion, AM7021) in 4°C for 1 h, fixed in 4 % paraformaldehyde (Sigma, SI-P6148) in 0.1 M PB in 4°C for 6 h, and

cryoprotected with 30% sucrose (J.T.-Baker, JT-4097-04) in 0.1 M PB in 4°C overnight. Fixed brains were placed into a cryomold, embedded with O.C.T. compound (Surgipath, FSC 22) and frozen in pre-chilled isopentane (Sigma, AL-M32631) bath cooled with dry ice. Frozen brains were sectioned into 7 μm thickness with microtome (Leica, CM3050 S) and collected on membrane slide (Zeiss, 415190-7041-000). Sections were dried and dehydrated by performing one 30 second-dips in 50% ethanol (J.T. Baker, JT-8006-05) in DEPC (Sigma, D5758)-treated water, one 10 second-dip in 95% ethanol in DEPC-treated water, two 30 second-dips in 100% ethanol and two 2 minute-dips in 100% xylene (J.T. Baker, JT-9490-03). Red fluorescent positive cells in the primary visual cortex were captured using a laser capture microdissection system (Zeiss, PALM MicroBeam) and collected into AdhesiveCap 500 opaque PCR tubes (Zeiss, 415190-9201-000). For the first two hours, tubes were replaced every 30 minutes. After that, tubes were replaced every 400 captured cells. 6000 cells were captured before further process.

Visual cortical RNA isolation, library preparation and sequencing

Total RNA was extracted by NucleoSpin miRNA (Macherey-Nagel, 740971), quantified by a ND-1000 spectrophotometer (Nanodrop

Technology) and qualitated by Bioanalyzer 2100 (Agilent Technology) with RNA 6000 LabChip Kit (Agilent Technology). Purified RNA was amplified and prepared for sequencing by TruSeq RNA Sample Prep Kit v2 (Illumina, RS-122-2001). Libraries were sequenced on an Illumina HiSeq2000 (100 PE bp) at the Welgene Biotech company and generated 6 Gb of data per sample.

Excitatory neuronal RNA isolation, library preparation and sequencing

Total RNA was extracted by totalRNA FFPE XS kit (Macherey-Nagel, REF740969) with minor modifications. In short, cells in the cap were covered with lysis buffer, incubated at 56°C for 15 min and the spun down. Extracted RNA was bound onto the column with binding buffer and genomic DNA was digested by 15 minute rDNase incubation. After two times of washes, RNA was eluted by 10 ml of RNase-free water. rRNA was removed from purified RNA by Ribo-Zero Magnetic Gold Kit (Epicentre, MRZG126). After that, purified RNA was amplified by SMARTer Ultra Low RNA Kit for Illumina Sequencing (Clontech, 634936) and followed by Low Input Library Prep Kit (Clontech, 634947). Libraries were sequenced on an Illumina HiSeq2000 (100 PE bp) at the Welgene Biotech company and generated 6 Gb of data per sample. RNA and DNA were

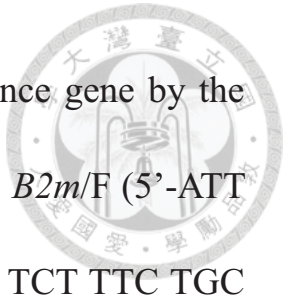
quantified and qualitated by Bioanalyzer 2100 (Agilent Technologies) with RNA 6000 Pico LabChip (Agilent, 5065-4401) and High Sensitivity DNA analysis Chip (Agilent, 5067-4672) respectively.

RNA-seq analysis

Initially, the sequences generated went through a filtering process to obtain qualified reads. ConDeTri was implemented to trim or remove the reads according to the quality score. Qualified reads after filtering low-quality data were analyzed using TopHat/Cufflinks for gene expression estimation. The gene expression level was calculated as FPKM (Fragments Per Kilobase of transcript per Million mapped reads). For differential expression analysis, CummeRbund was employed to perform statistical analyses of gene expression profiles. The reference genome and gene annotations were retrieved from Ensembl database. Genome Analysis Toolkit (GATK) was used for SNP finding and genotype determination. MMSEQ was applied for Allelic-Specific-Expression (ASE) analysis.

Real-time qPCR analysis

qPCR was run in StepOnePlus Real Time PCR Systems (Applied Biosystem) with Power SYBR® Green RNA-to-CT 1-Step Kit (Applied Biosystem, 4389986). All samples were triplicated and their relative



expression values were determined using *B2m* as a reference gene by the comparative Ct method (2-ΔΔCt). Primers were as follows: *B2m*/F (5'-ATT CAC CCC CAC TGA GAC TG-3'), *B2m*/R (5'-GCT ATT TCT TTC TGC GTG CAT-3'), *Arc*/F (5'-GGC TGG AGC CTA CAG AGC-3'), *Arc*/R (5'-GCC AGG CAC CTC CTC TCT-3'), *Bdnf*/F (5'- GCG CCC ATG AAA GAA GTA AA-3'), *Bdnf*/R (5'- TCG TCA GAC CTC TCG AAC CT-3'), *Camk2a*/F (5'-GAG CAG CAG GCA TGG TTT-3'), *Camk2a*/R (5'- GGA TAC CCA ACC AGC AAG AT-3'), *Gfap*/F (5'-CAC CTA CAG GAA ATT GCT GGA GG-3'), *Gfap*/R (5'-CCA CGA TGT TCC TCT TGA GGT G-3'), *Gad2*/F (5'-GCT TCT GGT TTG TAC CTC CTA G-3') and *Gad2*/R (5'- CCT AAG GGT TGG TAG CTG AC-3').

Statistical analysis

All data are presented as mean ± s.e.m., with sample sizes (n) shown in figures or stated in the text. Statistical analyses were performed using SigmaPlot 11 (Systat Software). Normality tests (Shapiro-Wilk) and equal variance tests were run and passed ($P > 0.05$) before parametric statistical analyses were run.

Results

The Cre/loxP recombination system in transgenic mice expresses fluorescent protein specifically in the excitatory neurons of the mice brains

First of all, I needed to prove that two of our transgenic mice have the correct genotypes. That is, *Camk2a-iCre* mice expresses the iCre recombinase enzyme under the control of the *Camk2a* gene promoter. *Ai14* mice contain the sequences with a *loxP*-flanked STOP cassette preventing the transcription of the downstream red fluorescent protein variant (tdTomato). With the sequences of *Camk2a-iCre* mice (Figure 3a), I designed the primer which is specific for *iCre* and its product size is 381 bp (Figure 3c). For the genotyping of *Ai14* mice, I used the primer from JAX lab (Figure 3b) and yields wide type band (196 bp) and mutant band (297 bp) (Figure 3d).

Next, I mated these two transgenic mice. With the genotype containing both *Cre* and *loxP* sequence, offspring would have the deleted-STOP cassette and red fluorescent protein would be only expressed under the control of *Camk2a* gene promoter. CAMK2A is only expressed in excitatory neurons. In order to confirm our red-fluorescent-positive cells

were *Camk2a-iCre*-positive neurons, I used *CAMK2A* antibody to see the co-localization between green fluorescent signals (against to CAMK2A-antibody) with red fluorescent signals from the expression of *Camk2a-iCre*. (Figure 4). I then wanted to determine the expression profile of my *Camk2a-icre::Ai14* mice, strictly speaking, their developmental profile and distribution pattern. Informed by Figure 5, I can observe that *Camk2a* starts to be expressed from P28 and its fluorescent signal were detected most abundantly in the cortex, hippocampus, resembling the pattern of endogenous *Camk2a* gene.

Sample preparations for laser capture microdissection

Next, I intended to optimize the protocol of dissecting my samples in 7 μm thickness and capture the fluorescence-labeled cells with laser capture microdissection. Since I wanted to conduct RNA-seq after collecting our cells, the quality of tissue is the critical determinant factor to all of it. At the first place, I dissected the brains without any treatment. Figure 6a showed the results after I sectioned fresh brain in 7 μm, and couldn't detect any fluorescent signal even the genotype is right ($icre^{+/-}$::$td^{+/-}$). For modified method, I hypothesized that fluorescent protein must be fixed before dissecting our samples. So I tried the second and third methods which mice

brains were fixed by paraformaldehyde (PFA) through perfusion and few drops of incubation. Results from PFA-perfusion (Figure 6b), I was able to detect some fluorescent-labeled cells but the background signal is so strong that we could not observe which cells are fluorescent-labeled. For the third modification (Figure 6c), the images showed no signals. Herein, fluorescent protein must be fixed before dissection and perfusion is not an appropriate way to fix the brain. To achieve the purposes of fixing tissue as well as possible without enhancing the background, I immersed the brain in the PFA immediately after removing it from the skull for one overnight. Cryoprotected with 30% sucrose for two overnight, I eventually could detect the fluorescent-labeled cells through 7 um dissection.

Since PFA-treatment could destroy the quality and quantity of RNA[36], so I run some tests on PFA-treated tissues before extracting RNA from LCM-collecting-cells. Fresh brains were used as positive controls (Figure 7a) and yield approximately about 621 ng/mg in RNA quantity, however, the amount of RNA decreased as immersed time increased. (Figure 7b) PFA-treatment dramatically influenced RNA quantity, and caused no differences after 1 hr of immersing. To further enhance RNA quantity, I added the step of de-crosslinking during the time of extracting RNA. From

Figure 7c, I can tell that the de-crosslinking treatment maximize the amount of RNA. Putting the abundant RNA apart, fluorescent signal could only be detected after 3 h of immersing and appeared vividly after 6 h (Figure 7d). That is, considering both the quantity of RNA as well as the incubation time of PFA for the brain, I am now capable of extracting RNA from PFA-treated samples. On the other hand, quality of RNA was dropped from the integrity of RIM number 9 to 2.2. (Figure 7e)

Steps further, same protocol had been reproduced on cellular level. Figure 8a1 illustrated how I circulated the cells under fluorescent microscope and captured cells by laser pulse to the cap. Adhered to the cap (Figure 8a2), cells then be gathered and stored for RNA extraction. Given the reason that I gathered the cells on visual cortex and restricted in layer 2、3, I confirmed our results by using Nissl staining method which could tell us whether I meet the criteria or not. Extracting RNA from the cells I collected which according to the protocol described earlier, I then used Bioanalyer 2100 to check the quality and quantity of our RNA. Roughly speaking, approximately about the amount of 200 pg for 100 cells (Figure 8d). In order to assess the accuracy, I designed the cell type-specific primer to confirm our results again (Figure 8e), which only the excitatory neuron

marker (*Camk2a*) showed up the product instead of interneuron marker (*Gad2*) or Astrocyte marker (*Gfap*).

RNA amplification and library build-up for RNA-seq

Acknowledging that the amounts of RNA were not sufficient for RNA-seq, I then developed some strategies to conquer the problems. First, I assumed that more cells to be collect were have more RNAs, I then set the goal to ultimately gathering 6000 *Camk2a*-positive neurons. Second, I used amplification-kit, explicitly designed for amplifying PFA-treated sheared RNA, and suited for further conducting RNA-seq Figure 9 demonstrated our captured cell RNAs are sufficient for further RNA-seq Of the first strategy, the very beginning 500 cells had the concentration better than second 2000 cells (50 pg/ul) and the total amount of RNA is enough for amplification (all three elute in RNase-free water in total volume of 30 ul). Hence, I thought it may not necessary to meet the number of 6000 because of quantity achieved as well as decaying of RNA through time. I built up library after amplifying RNA with Smarter kit, I used TS HS D1000 to check the quality and quantity of cDNA, which all reached the criteria for further RNA-seq. (total amount > 4ug, length> 260 bp)

RNA-seq of the visual cortex

In order to compare between our cellular-level sequencing, I first sent out total RNA which extracted from whole visual cortex as a control. Taking the advantage of strain-specific SNPs, I am able to identify whether transcripts express from mother or father (monoallelic expression) or both (biallelic expression). To determine the cutoff which help me to screen whether the genes are imprinted or not, I first made scattering diagrams to see the expression pattern when I take the gene's expression SNP as a call comparing between the amount of paternal with maternal expression. Figure 10 shows the results. Of one gene, each availably mapping SNPs were taking into account, and the amount of maternally expression SNPs were divided with the paternally expression. Take log of the number above, I constructed the pattern demonstrated in the figure 10. No matter from C::B6 mating or B6::C mating, I can see three cohort groups: the middle scattering groups were the genes equally expression from paternal or maternal (biallelic expression), and the upper right as well as lower left stands for the genes with unequally expression (monoallelic expression). Statistically speaking, the upper right groups and lower left had the numerical value beyond two standard deviations (abbreviated as 2SD group), which were statistically significant difference. As the reason above,

I believed that imprinted genes should among these groups because of the exclusively expression nature of imprinted genes. Consequently, I correlated the 2 x S.D. groups of C::B6 mating with B6::C mating, to see the matched genes as a screening strategy. Using this procedure, I am able to identified 8 genes, three maternally expression genes and five paternally expression genes (Figure 11a). However, imprinted genes which have been identified nowadays were the number approximately of 120[37], far more than the number I got. Realizing the cutoff I made might too strict, I then lower down the acceptable range to 1 standard deviation. Although the genes been screening out up from 8 to 13, it still didn't make the fractions of 120 genes.

Of the strategies failed to screen down the matched number of well-known imprinted genes, I then started to wonder what is the appropriate cutoff point. Realizing that there were no currently-accepted criteria for screening, I then mapping those well-confirmed imprinted genes back to our sequence data. By doing so, I might have the lowest point to be the cutoff point for imprinted genes. Figure 12a and 12b demonstrated that the results which I paralleled the lists of imprinted genes back to my RNA-Seq database. Because some of the genes only-confirmed in the embryonic stage, the

most likely criteria should be ±0.5 since the genes at this point acclaimed to be imprinted in adult. Figure 13 were the genes filtered in the criteria more than ±0.5 which yields the total of 193 genes. Although the total counts are much more reasonable, I still need to confirm the imprinting status of these genes again.

Environmental effect on the parent-of-origin-specific expression in the mouse visual cortex

Because I wanted to investigate whether the condition of dark rearing could influence the imprinting status or not, I first put our pregnant mice in the whole dark rearing setting. Those litters were born under total dark environment. On the day of postnatal 28, I then took out the mice brain, and dissected out the visual cortex. According to previous studies[38], several genes are down-regulated under dark-rearing condition. I chose *Arc* and *Bdnf* as my positive control genes to prove my dark-rearing system could work. Comparing to the light rearing samples, my dark rearing samples do really display significant decline (Figure 14b). Thus, I then wanted to do genome-wide RNA-Seq on my dark-rearing visual cortex before processing imprinting analysis on excitatory neurons of visual cortex. I predicted that the imprinting status could be relaxed or switched. I chosed the samples with the most predominantly changed of *Arc* and *Bdnf* expression (F1 &

R2.) (Figure 14c) for RNA-Seq.

Discussion

Briefly, I would like to divide my thesis into two parts. The first part is that I wanted to profile the parent-of-origin-specific expression in the excitatory neurons. In order to achieve that goal, I need to build up a platform to conduct single-cell RNA sequencing in the excitatory neurons of mouse visual cortex and identify potential imprinted genes. My initial hurdle for this part is to detect fluorescent signals within 7 μm thick sections. Surprisingly, I didn't detect any fluorescent signal from fresh 7 μm thick sections. That means fluorescent protein can't kept within 7 μm thick sections without further fixation. Since we will profile parent-of-origin-specific expression in the human brain and most human brain were fixed already, it would be very useful if we develop a good platform to get good quality and quantity RNAs from fixed cells.

For the second part, I wanted to see whether environmental factors could influence the status of genomic imprinting. I chose mouse visual cortex as an experience-regulated system to address my second question. The maturation of mouse visual cortex is between P22 and P28 under light guidance[39,40]. And since that imprinted genes have been proved to participate in development process, we hope to see imprinted status being

destroyed due to irregular rearing condition.

In order to compare with our cellular-resolution sequencing data, I first sent out whole visual cortex for RNA-seq. as a control. Since there is no commonly agreeable criteria for determining imprinted genes from RNA-seq. dataset, I tried different criteria and tentatively chose -$\log_{10}$ (maternally FPKM/ Paternally FPKM) > ±0.5- as a cutoff for determining imprinted genes. Based on this standard, I narrowed down to 193 genes as imprinted gene candidates. I am under the way to validate these candidates by Sanger sequence and hope to find novel imprinted genes. There's also other interesting findings could be discussed regarding to monoallelic gene expression. From our data, we could see the genes monoallelic expression further categorize into three groups: stochastic expression, strain-specific expression, and imprinted genes. Of the genes being biallelic expression, they could benefit on the compensate dosage of homologous genes if one of them loss their function. But why we risk developing a way randomly silenced one of the allele when we speaking about stochastic genes especially they formed the majority percentage in our data? It's still need further investigation. Of the strain-specific expression, some other groups[41] had already found same phenomenon and may count the

strain-specific-behavior differences.

At last, the validation of our dark-rearing system was confirmed by *Arc* and *Bdnf* expression level. Therefore, I sent out two of our dark rearing samples from C::B6 and B6::C mating pairs. I hope these environmental factors could influence the imprinted status in mouse visual cortex.

Because of the exogenous nature of tissue, cellular-resolution RNA-Seq had been a way narrowing down the confounding factors especially there's a lot of ways targeting the specific cells we want currently. But there are also many obstacles needed to be dealt with such as RNA degradation, contamination problems, and the small amounts of RNA in one cell. Our group developed some strategies to conquer the problems above, and might the first group genome-wide analysis imprinted status by using cellular-resolution RNA-Seq. Although I still have not got the results yet, I currently now proved that our system could yield applicable library to conduct RNA-Seq and hope to find novel imprinted genes or identify cell type-specific imprinted genes.

There's a lot of works which haven't been done yet. The first thing is that I need to double confirm the potential imprinted genes which been screened out. I am using Sanger sequencing now to complete the work and hope to

find the cutoff line for screening. The second work is to analysis my dark rearing samples, using same strategies mentioned above, we want to find some genes being switched, turned-on, or turned-off of their imprinted status due to environmental manipulation. Lastly, analyzing my RNA-Seq data with cellular-resolution could be my future works since the samples have been send for sequencing.

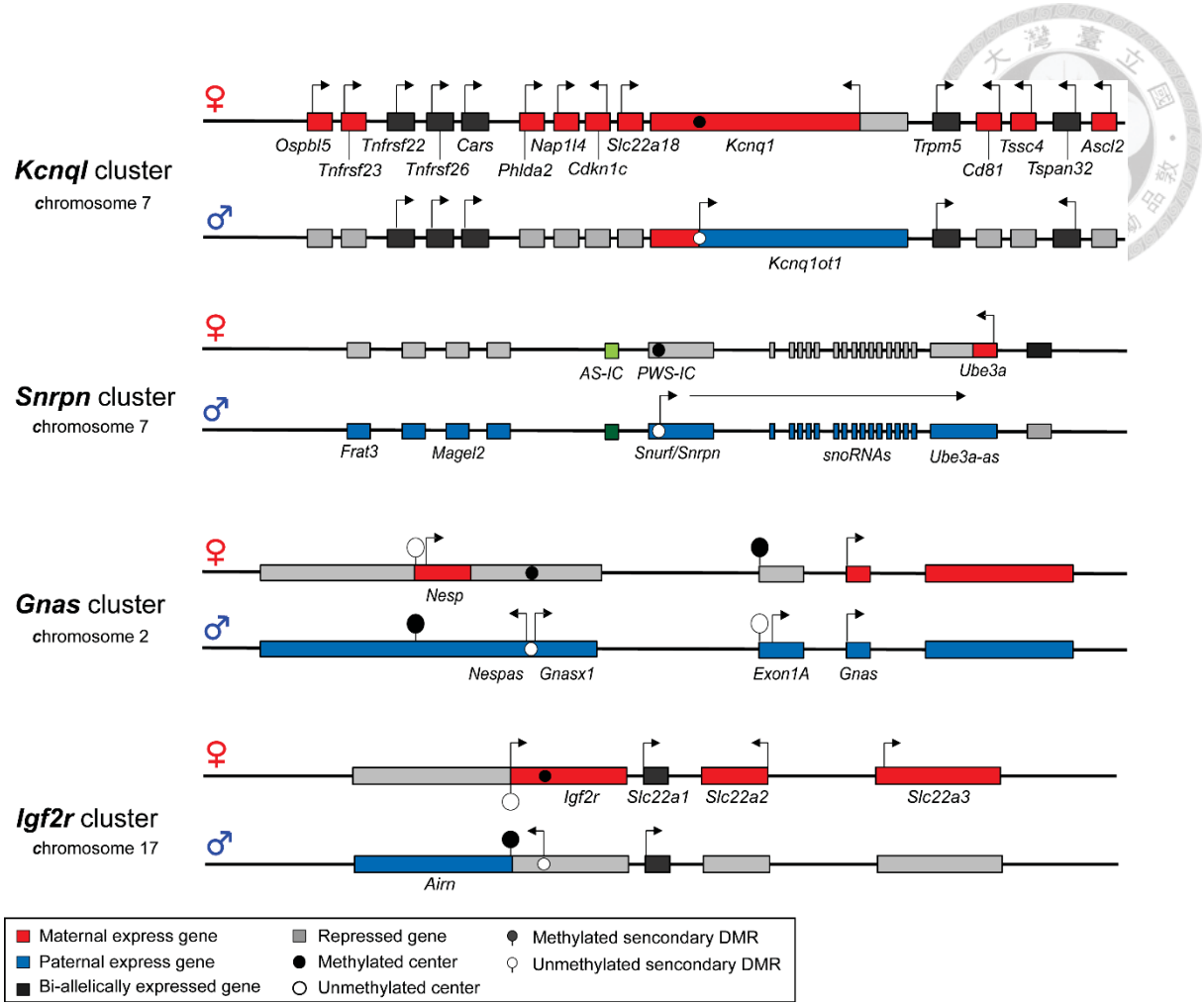


Figure 1. Four well-known imprinted gene clusters. (adapted from Feguson-Smith, A.C., ***Nature Review Genetics***, 2011)

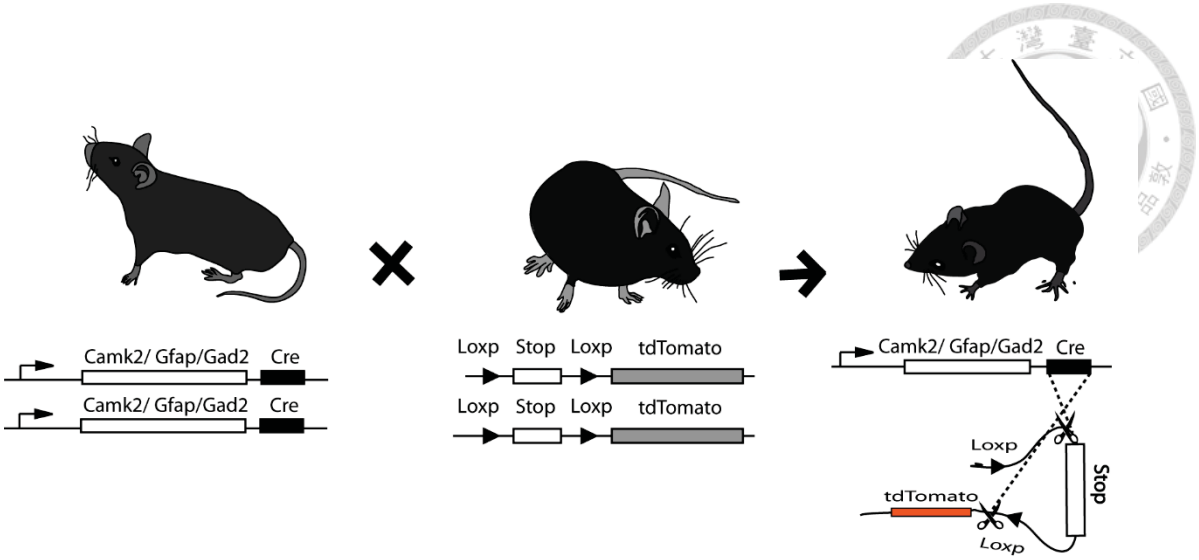


Figure 2. Cell type-specific Cre/lox P recombination system. Cell type-specific Cre line mice was mated with *Ai14* (transgenic mice containing a loxP-flanked STOP cassette before fluorescent protein-tdTomato) and yields the mice with fluorescent protein under the control of cell type-specific promoter.

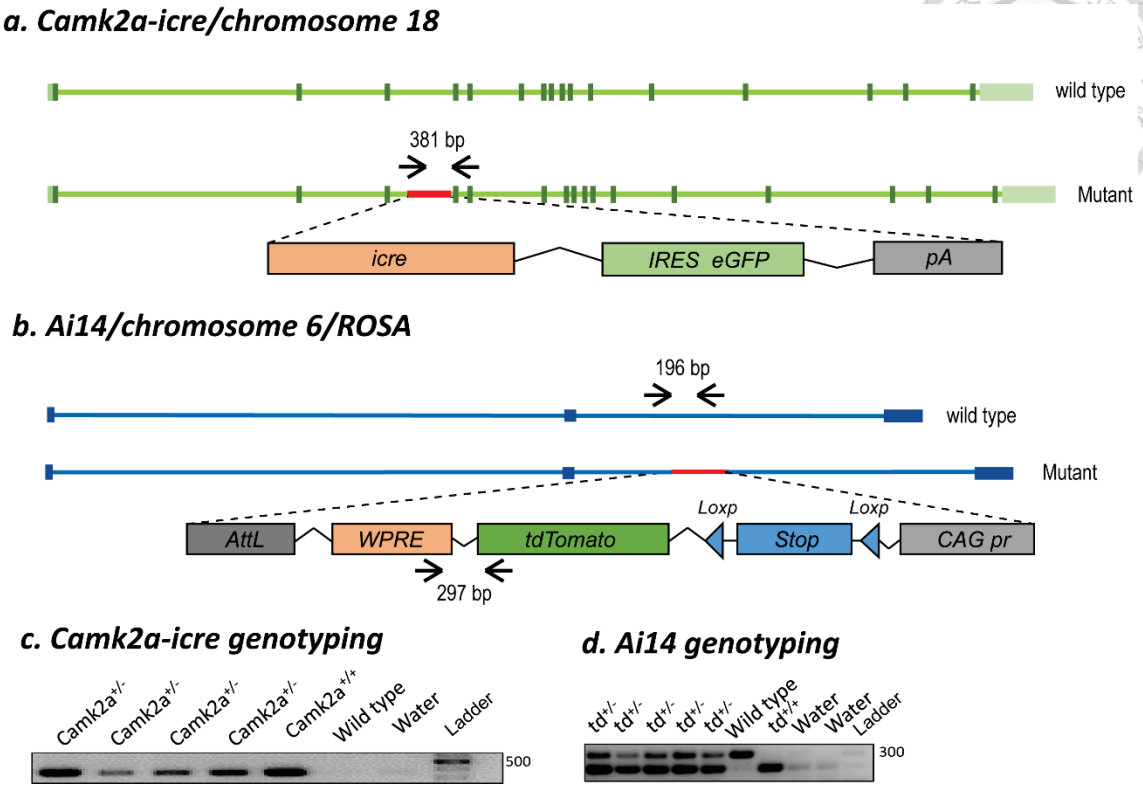


Figure 3. Genomic structure and genotyping of *Camk2a-iCre* and *Ai14* mice. a, Genomic structure of *Camk2a-iCre* mice. The double headed arrows stand for the locations of forward and reverse primer respectively. **b**, Genomic structure of *Ai14* mice. **c, d**, Genotyping results of *Camk2a-iCre*, *Ai14*, and their litters. $Camk2a^{+/-}$ means heterozygous mutant. $td^{+/-}$ and $td^{+/+}$ stands for *Ai14* heterozygous and homozygous mutant, respectively.

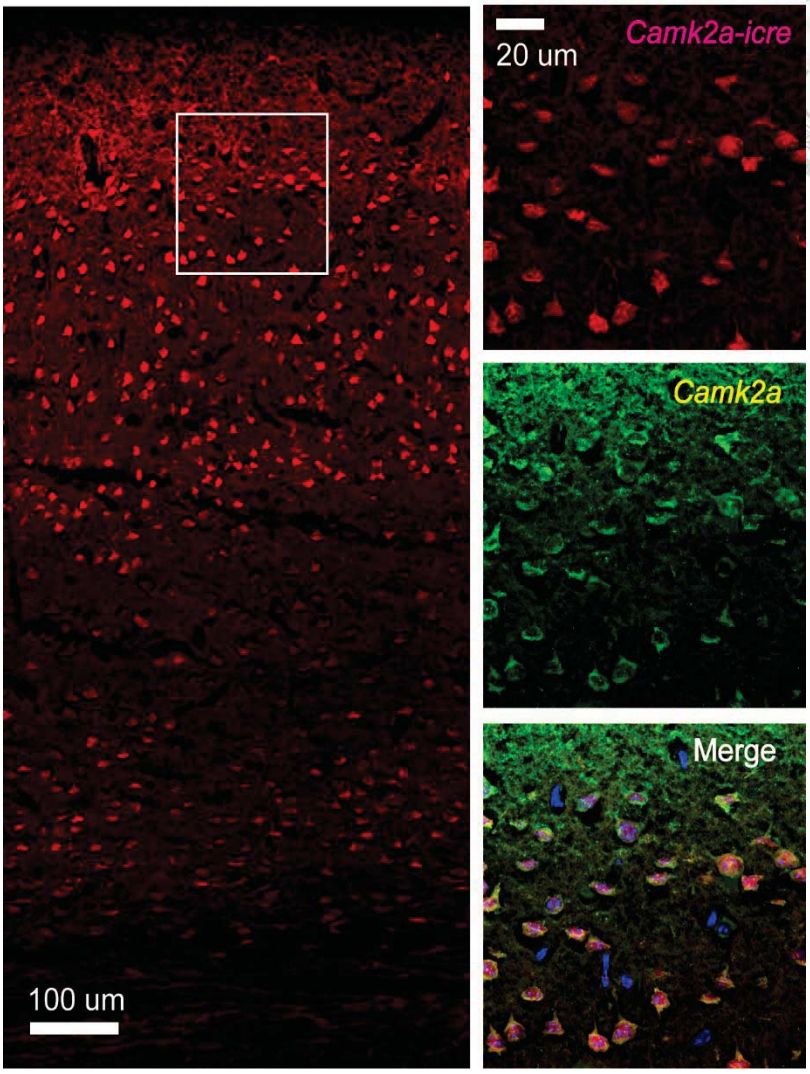


Figure 4. *Camk2a-iCre::Ai14* mice specifically display fluorescent protein in the excitatory neurons of the visual cortex. Left panel shows the distribution pattern of red fluorescent signal (tdTomato fluorescent protein) in the visual cortex of *Camk2a::Ai14* mice at the postnatal day 28. Right enlarged panels show the co-localization of tdTomato fluorescent protein (red) and endogenous CAMK2A protein (yellow), an excitatory neuron marker. Sections were counterstained with the nuclear marker, DAPI (blue). Scale bar = 100 μm (left panel), 20 μm (right panel).

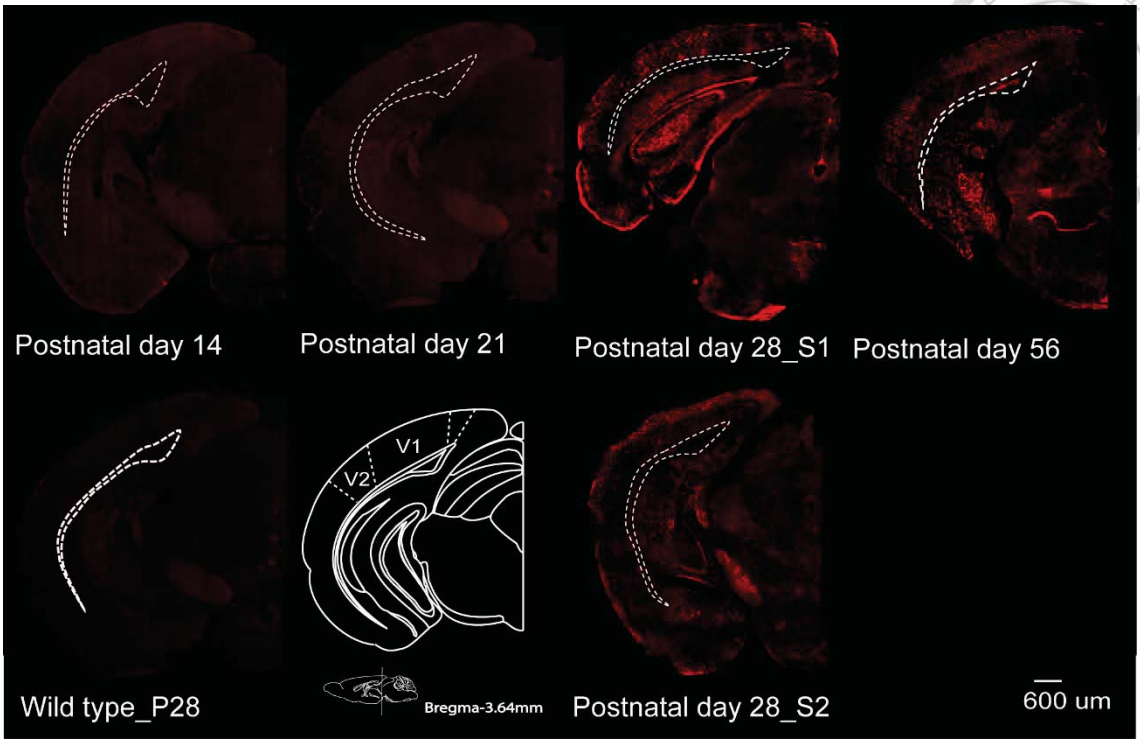


Figure 5. Postnatal developmental profile of *Camk2a-iCre* expression in the brain of *Camk2a-iCre::Ai14* mice. Postnatal day at 14 (P14), 21 (P21), 28 (P28) and 56 (P56) were chosen for the developmental profile of *Camk2a-iCre* expression (red). Two brain samples at P28 (S1, sample1/S2, sample 2) were used and one wild type brain at P28 was used as a negative control. All samples were processed with the same protocol and the thickness of brain section is 7 μm. Scale bar = 600 μm.

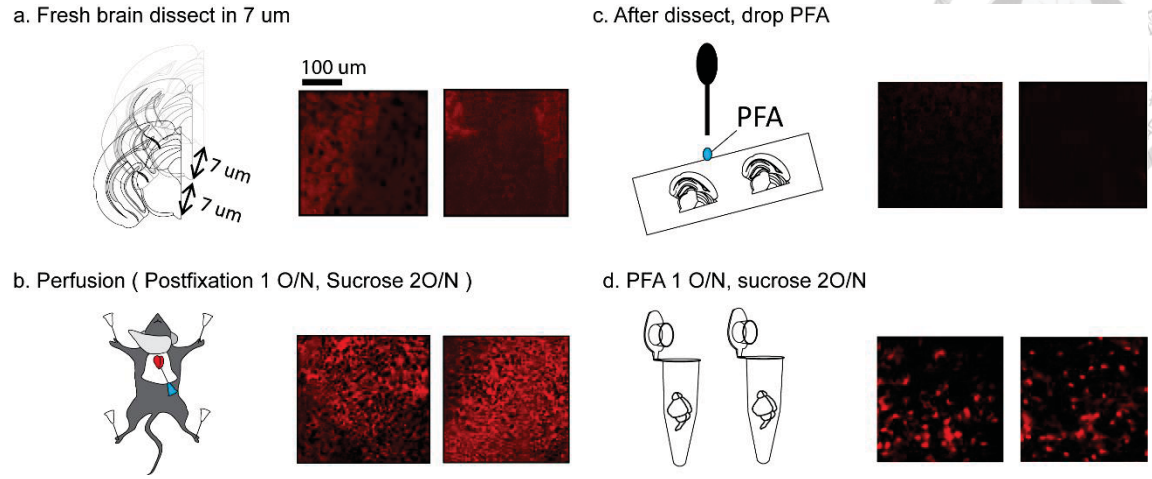


Figure 6. Different preparations of *Camk2a-iCre::Ai14* mice brains for laser capture microdissection. a, Fresh mouse brain was sectioned into 7 μm thickness without any further treatment. **b,** Mice was perfused with 4% paraformaldehyde (PFA), postfixed for one overnight and saturated with 30% sucrose for 2 overnight. **c,** Fresh mouse brain was sectioned into 7 μm and fixed with a few drops of PFA for 30 mins. **d,** Fresh brain was immersed in PFA for 1 overnight and saturated with 30% sucrose for two overnight.

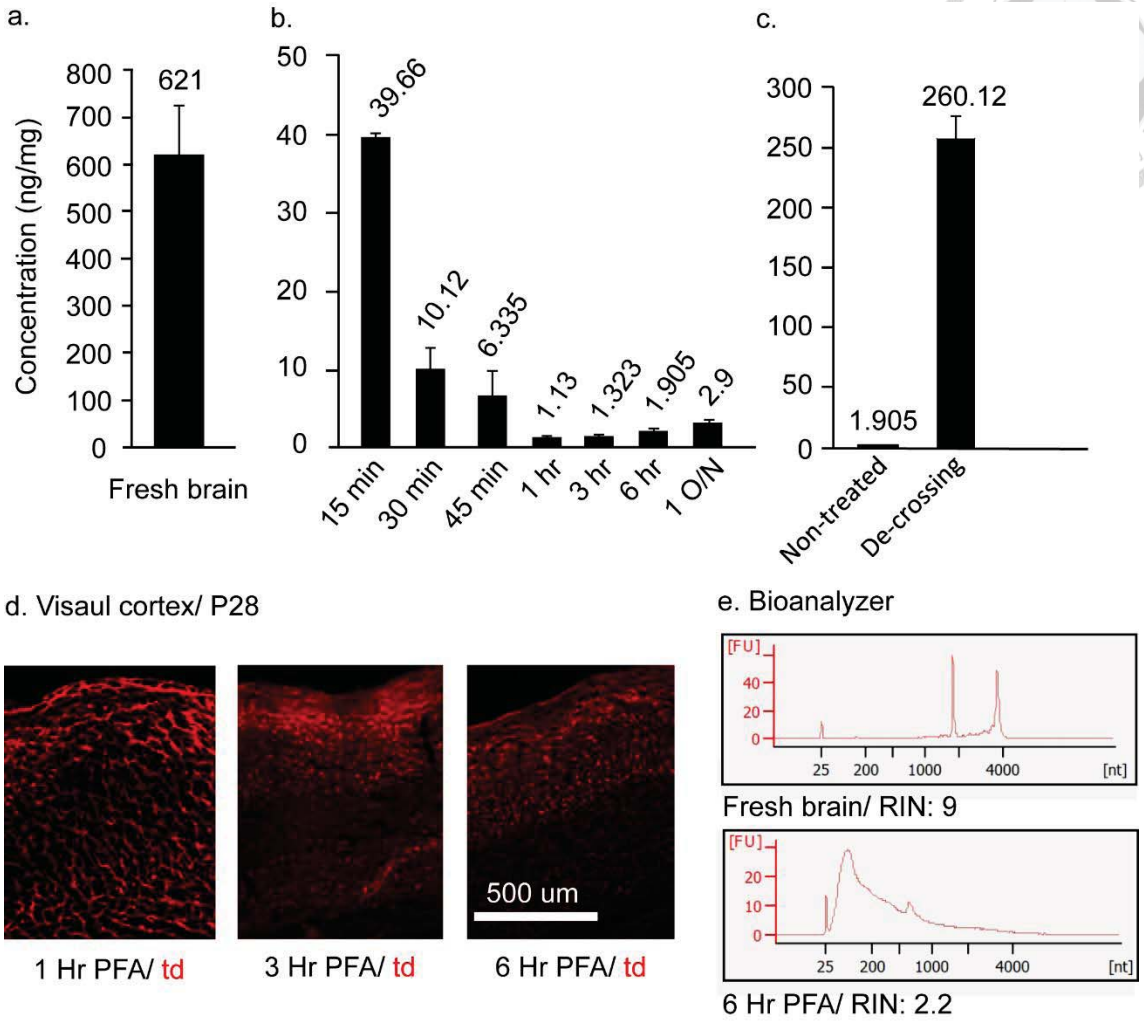


Figure 7. Quality and Quantity of RNAs from PFA-treated brain samples. RNA quantity of fresh brains **(a,)**, PFA-fixed brains with different fixation time **(b,)** and PFA-fixed brain with or without de-crosslinking treatment **(c,)** were measured. **d,** Fluorescent signals of tdTomato protein were observed at three fixation time in the mouse visual cortex. Scale bar = 500 μm. **e,** RNA quality was determined by Bioanalyzer 2100 and indicated as the RIN number.

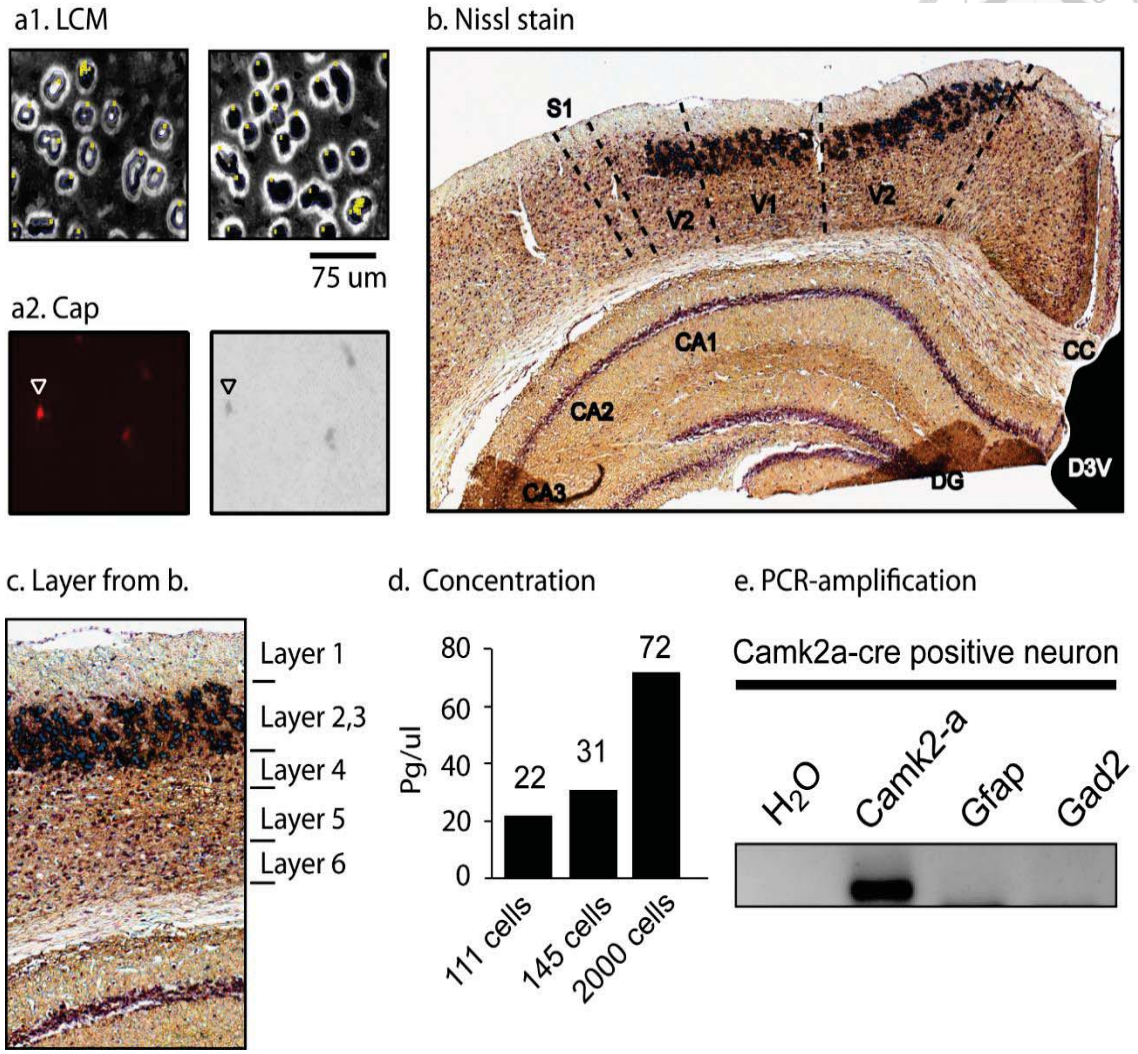


Figure 8. The overall process of laser capture microdissection in the mouse brain. **a1,** Fluorescent-positive cells were circulated by laser beam (left panel) and captured by laser pulse (right panel). **a2,** Captured cells on the cap were checked by fluorescent (left) and white (right) light source. **b & c,** Nissl stained coronal sections after LCM confirmed our captured cells were primarily located on layer 2 & 3 of the visual cortex. **d,** RNA concentrations were quantified by Bioanalyzer 2100 and their corresponding cell number was indicated. **e,** The purity of captured cells was confirmed by PCR with cell-type specific primers. *Camk2a* primer is specific for excitatory neuron, *Gfap* primer is specific for astrocyte and *Gad2* primer is specific for interneuron.

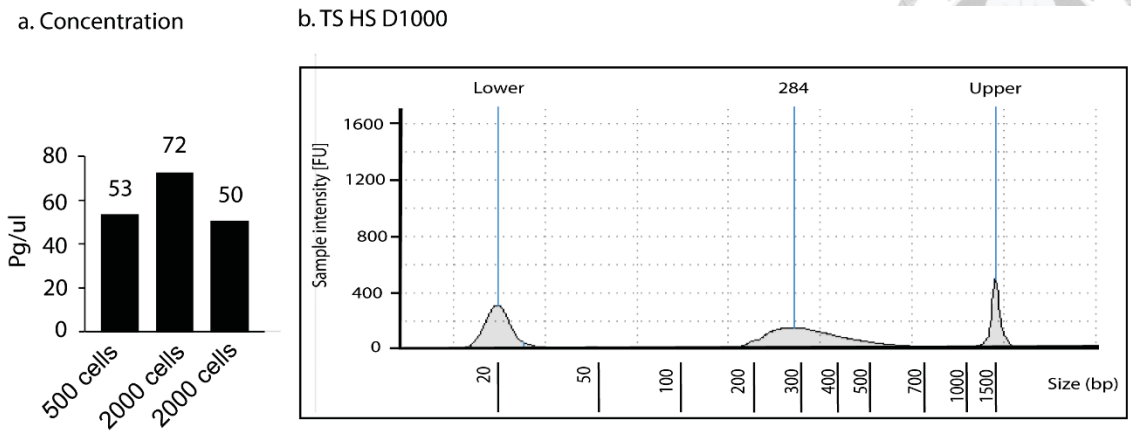


Figure 9. RNA concentration of captured excitatory neurons and their further library buildup for single-cell-type RNA-sequencing. a, RNA concentration of captured cells were measured by Bioanalyzer 2100 and their corresponding cell number was indicated. **b,** After RNA amplification and cDNA library buildup, its quality and quantity was determined by TS HS D1000. Our sample yielded 284 bp in average and 7 ng in total, which meet the criteria of RNA-seq..

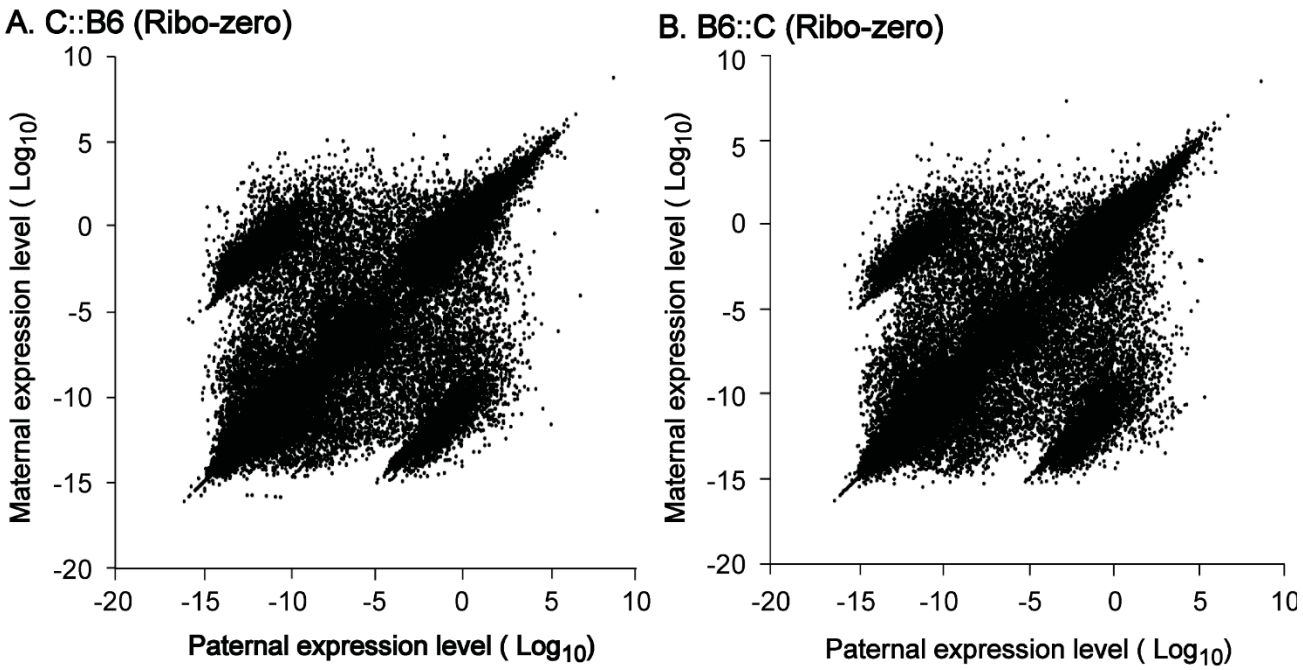


Figure 10. Parental expression levels in the offspring's visual cortex from C::B6 and B6::C mating pairs. Scattering diagram of C::B6 pair **(a,)** and B6::C mating pair **(b,)**. Each dot stands for parental expression level of a gene.

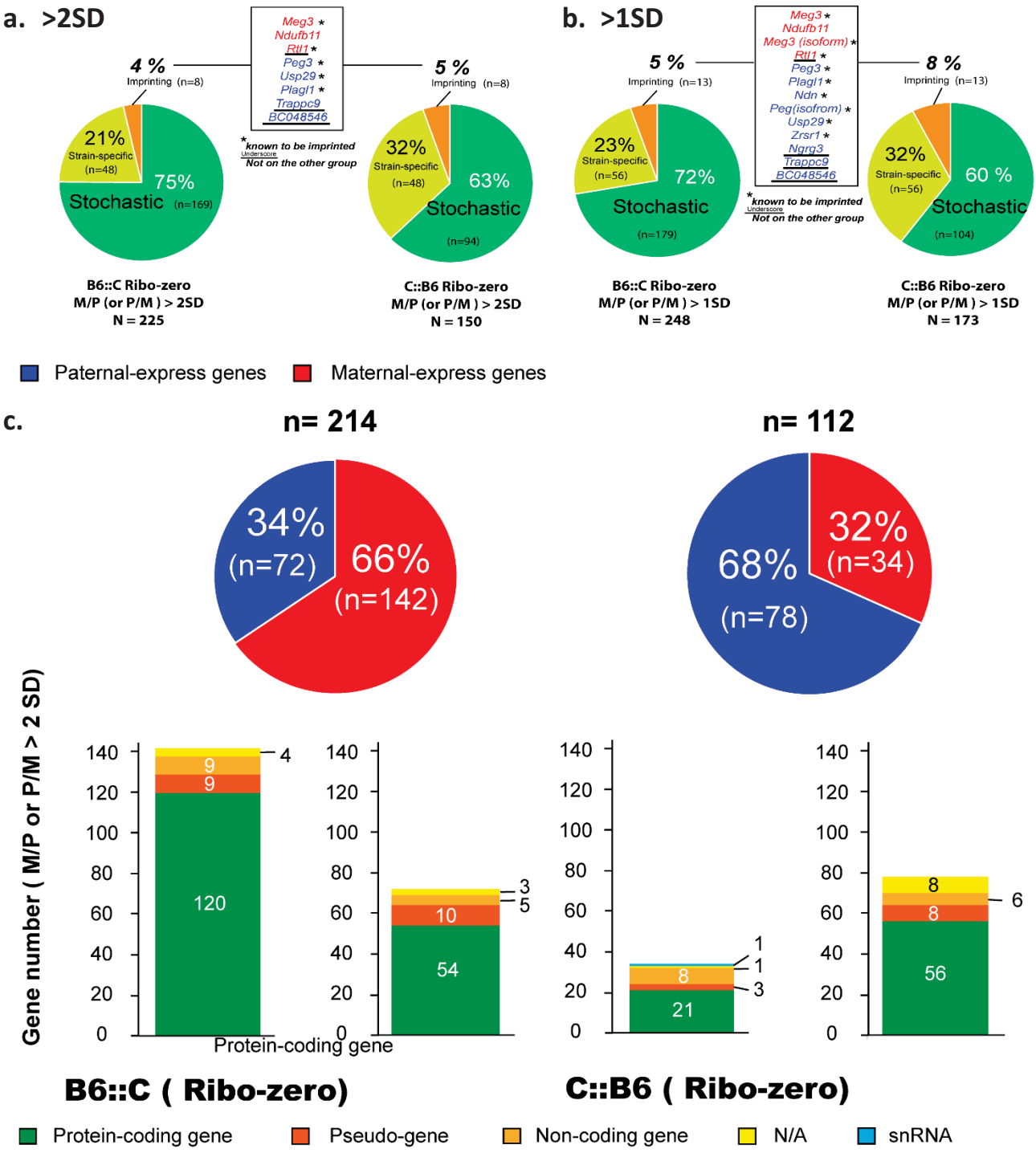


Figure 11. Monoallelically-expressed genes were identified through different cutoffs. The genes indicated monoallelic expression pattern from C::B6 and B6::C mating pairs. The genes whose value of $\log_{10}$(maternal expression level/paternal expression level) is above 2 standard deviation (S.D.) **(a,)** or 1 S.D. **(b,)** were selected. **c**, Classification of monoallelically-expressed genes from **a.**

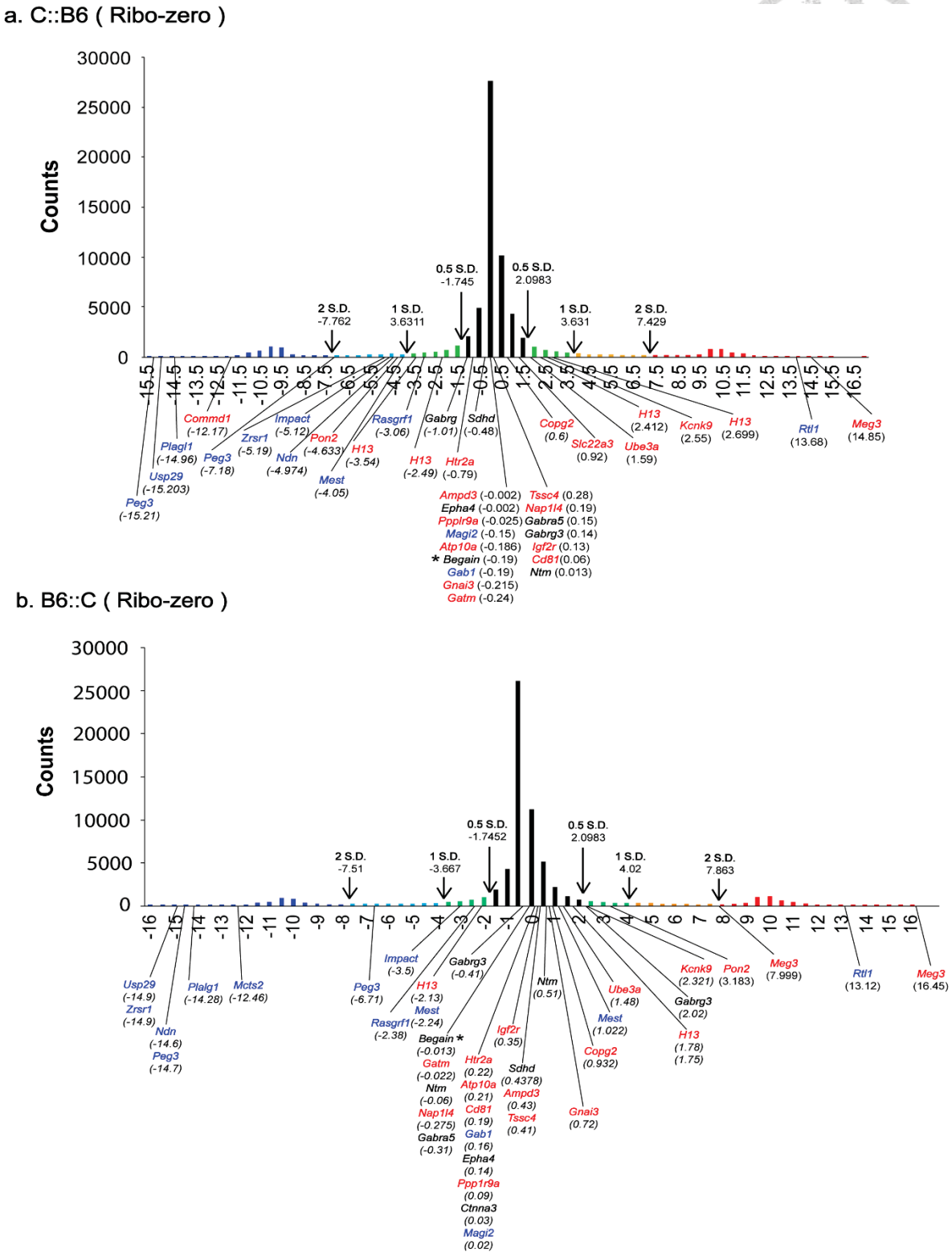


Figure 12. Distribution pattern of known imprinted genes in our C::B6 and B6::C pair dataset. Known imprinted genes in our C::B6- **(a,)** and B6::C- **(b,)** mating pair dataset. Maternally expressed imprinted genes were red color and paternally expressed imprinted genes were blue color. The value of $\log_{10}$(maternal expression level/paternal expression level) was labeled in parenthesis.

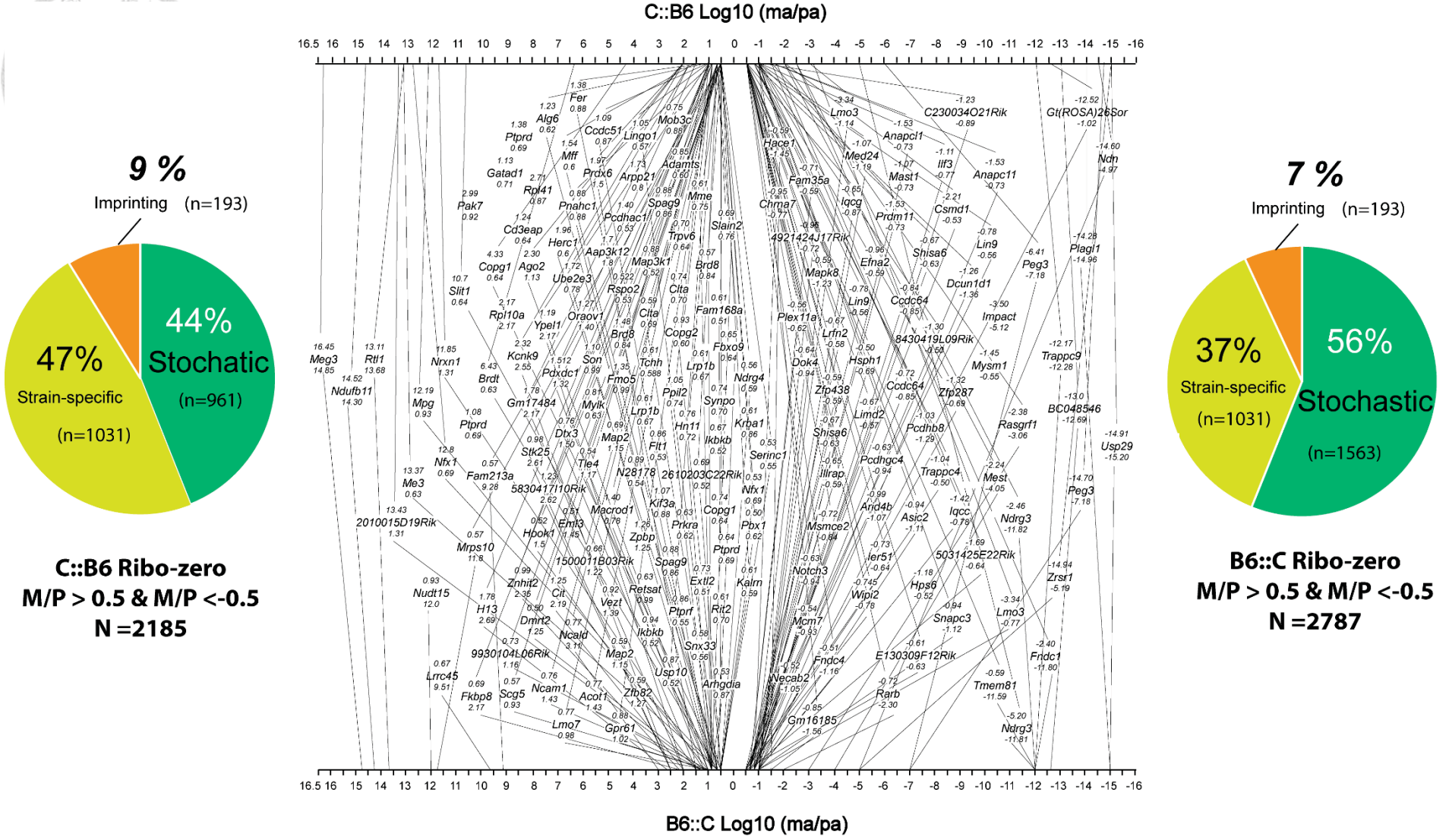


Figure 13. Candidate imprinted genes. One hundred and ninety-three candidate imprinted genes were screened by the cutoff ±0.5.

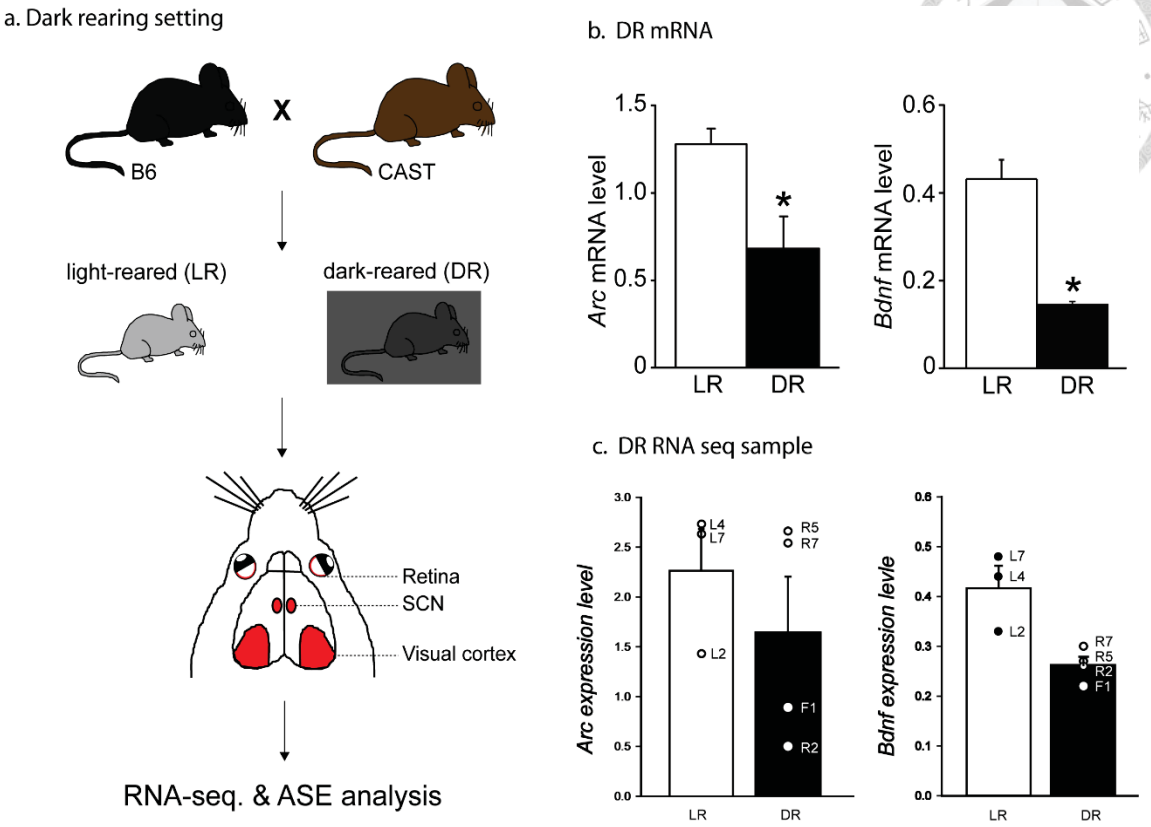


Figure 14. Identification of dynamic imprintome in the mouse visual cortex. **a,** Experimental scheme for identifying the light experience-regulated imprintome in the mouse visual cortex. **b,** Expression levels of activity-regulated genes such as *Arc* and *Bdnf* are lower in the dark-reared (DR) visual cortex than in its light-reared (LR) counterpart. **c**, *Arc* and *Bdnf* expression levels were measured in two DR samples (F1: C::B6 mating; R2: B6:C) before we sent them out for RNA-sequence.

References

1 Deng, Q., Ramskold, D., Reinius, B. & Sandberg, R. Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. *Science* **343**, 193-196, doi:10.1126/science.1245316 (2014).

2 Luedi, P. P. *et al.* Computational and experimental identification of novel human imprinted genes. *Genome research* **17**, 1723-1730, doi:10.1101/gr.6584707 (2007).

3 Reik, W. & Walter, J. Genomic imprinting: parental influence on the genome. *Nature reviews. Genetics* **2**, 21-32, doi:10.1038/35047554 (2001).

4 Wilkinson, L. S., Davies, W. & Isles, A. R. Genomic imprinting effects on brain development and function. *Nat Rev Neurosci* **8**, 832-843, doi:10.1038/nrn2235 (2007).

5 Luedi, P. P., Hartemink, A. J. & Jirtle, R. L. Genome-wide prediction of imprinted murine genes. *Genome research* **15**, 875-884, doi:Doi 10.1101/Gr.3303505 (2005).

6 Casanova, E. *et al.* A CamKIIalpha iCre BAC allows brain-specific gene inactivation. *Genesis* **31**, 37-42 (2001).

7 Lee, J. T. & Bartolomei, M. S. X-inactivation, imprinting, and long noncoding RNAs in health and disease. *Cell* **152**, 1308-1323, doi:10.1016/j.cell.2013.02.016 (2013).

8 Lepage, J. F. *et al.* Genomic imprinting effects of the X chromosome on brain morphology. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **33**, 8567-8574, doi:10.1523/JNEUROSCI.5810-12.2013 (2013).

9 Yamaguchi, S., Shen, L., Liu, Y., Sendler, D. & Zhang, Y. Role of Tet1 in erasure of genomic imprinting. *Nature* **504**, 460-464, doi:10.1038/nature12805 (2013).

10 Ferguson-Smith, A. C. Genomic imprinting: the emergence of an epigenetic paradigm. *Nature reviews. Genetics* **12**, 565-575, doi:10.1038/nrg3032 (2011).

11 Huang, H. S. *et al.* Topoisomerase inhibitors unsilence the dormant allele of Ube3a in neurons. *Nature* **481**, 185-189, doi:10.1038/nature10726 (2012).

12 King, I. F. *et al.* Topoisomerases facilitate transcription of long genes linked to autism. *Nature* **501**, 58-62, doi:10.1038/nature12504 (2013).

13 Wang, Y. *et al.* The mouse Murr1 gene is imprinted in the adult brain, presumably due to transcriptional interference by the antisense-oriented U2af1-rs1 gene. *Molecular and cellular biology* **24**, 270-279 (2004).

14 Yamasaki, K. *et al.* Neurons but not glial cells show reciprocal imprinting of sense and antisense transcripts of Ube3a. *Human molecular genetics* **12**, 837-847 (2003).

15 Ferron, S. R. *et al.* Postnatal loss of Dlk1 imprinting in stem cells and niche astrocytes regulates neurogenesis. *Nature* **475**, 381-385, doi:10.1038/nature10229 (2011).

16 Schmidt-Edelkraut, U., Hoffmann, A., Daniel, G. & Spengler, D. Zac1 regulates astroglial differentiation of neural stem cells through Socs3. *Stem Cells* **31**, 1621-1632, doi:10.1002/stem.1405 (2013).

17 Mardirossian, S., Rampon, C., Salvert, D., Fort, P. & Sarda, N. Impaired hippocampal plasticity and altered neurogenesis in adult Ube3a maternal deficient mouse model for Angelman syndrome. *Experimental neurology* **220**, 341-348, doi:10.1016/j.expneurol.2009.08.035 (2009).

18 Sato, M. & Stryker, M. P. Genomic imprinting of experience-dependent cortical plasticity by the ubiquitin ligase gene Ube3a. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 5611-5616, doi:10.1073/pnas.1001281107 (2010).

19 Wallace, M. L., Burette, A. C., Weinberg, R. J. & Philpot, B. D. Maternal loss of Ube3a produces an excitatory/inhibitory imbalance through neuron type-specific synaptic defects. *Neuron* **74**, 793-800, doi:10.1016/j.neuron.2012.03.036 (2012).

20 Yashiro, K. *et al.* Ube3a is required for experience-dependent maturation of the neocortex. *Nature neuroscience* **12**, 777-783, doi:10.1038/nn.2327 (2009).

21 Huang, H. S. *et al.* Snx14 regulates neuronal excitability, promotes synaptic transmission, and is imprinted in the brain of mice. *PLoS One* **9**, e98383, doi:10.1371/journal.pone.0098383 (2014).

22 Kishino, T., Lalande, M. & Wagstaff, J. UBE3A/E6-AP mutations cause Angelman syndrome. *Nature genetics* **15**, 70-73, doi:10.1038/ng0197-70 (1997).

23 Matsuura, T. *et al.* De novo truncating mutations in E6-AP ubiquitin-protein ligase gene (UBE3A) in Angelman syndrome. *Nature genetics* **15**, 74-77, doi:10.1038/ng0197-74 (1997).

24 Duker, A. L. *et al.* Paternally inherited microdeletion at 15q11.2 confirms a significant role for the SNORD116 C/D box snoRNA cluster in Prader-Willi syndrome. *European journal of human genetics : EJHG* **18**, 1196-1201, doi:10.1038/ejhg.2010.102 (2010).

25 Schanen, N. C. Epigenetics of autism spectrum disorders. *Human molecular genetics* **15 Spec No 2**, R138-150, doi:10.1093/hmg/ddl213 (2006).

26 Horike, S., Cai, S., Miyano, M., Cheng, J. F. & Kohwi-Shigematsu, T. Loss of silent-chromatin looping and impaired imprinting of DLX5 in Rett syndrome. *Nature genetics* **37**, 31-40, doi:10.1038/ng1491 (2005).

27 Moretti, P. & Zoghbi, H. Y. MeCP2 dysfunction in Rett syndrome and related disorders. *Current opinion in genetics & development* **16**, 276-281,

doi:10.1016/j.gde.2006.04.009 (2006).

28 Davies, W. *et al.* Xlr3b is a new imprinted candidate for X-linked parent-of-origin effects on cognitive function in mice. *Nature genetics* **37**, 625-629, doi:10.1038/ng1577 (2005).

29 Gregg, C. *et al.* High-resolution analysis of parent-of-origin allelic expression in the mouse brain. *Science* **329**, 643-648, doi:10.1126/science.1190830 (2010).

30 Gregg, C., Zhang, J., Butler, J. E., Haig, D. & Dulac, C. Sex-specific parent-of-origin allelic expression in the mouse brain. *Science* **329**, 682-685, doi:10.1126/science.1190831 (2010).

31 Wang, X. *et al.* Transcriptome-wide identification of novel imprinted genes in neonatal mouse brain. *PloS one* **3**, e3839, doi:10.1371/journal.pone.0003839 (2008).

32 DeVeale, B., van der Kooy, D. & Babak, T. Critical evaluation of imprinted gene expression by RNA-Seq: a new perspective. *PLoS genetics* **8**, e1002600, doi:10.1371/journal.pgen.1002600 (2012).

33 Ko, Y. *et al.* Cell type-specific genes show striking and distinct patterns of spatial expression in the mouse brain. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 3095-3100, doi:10.1073/pnas.1222897110 (2013).

34 Mellios, N. *et al.* miR-132, an experience-dependent microRNA, is essential for visual cortex plasticity. *Nature neuroscience* **14**, 1240-1242, doi:10.1038/nn.2909 (2011).

35 Morales, B., Choi, S. Y. & Kirkwood, A. Dark rearing alters the development of GABAergic transmission in visual cortex. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **22**, 8084-8090 (2002).

36 Goldsworthy, S. M., Stockton, P. S., Trempus, C. S., Foley, J. F. & Maronpot, R. R. Effects of fixation on RNA extraction and amplification from laser capture microdissected tissue. *Molecular carcinogenesis* **25**, 86-91 (1999).

37 Tunster, S. J., Jensen, A. B. & John, R. M. Imprinted genes in mouse placental development and the regulation of fetal energy stores. *Reproduction* **145**, R117-137, doi:10.1530/REP-12-0511 (2013).

38 Majdan, M. & Shatz, C. J. Effects of visual experience on activity-dependent gene regulation in cortex. *Nature neuroscience* **9**, 650-659, doi:10.1038/nn1674 (2006).

39 Lyckman, A. W. *et al.* Gene expression patterns in visual cortex during the critical period: synaptic stabilization and reversal by visual deprivation. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 9409-9414, doi:10.1073/pnas.0710172105 (2008).

40 Hensch, T. K. Critical period plasticity in local cortical circuits. *Nature reviews. Neuroscience* **6**, 877-888, doi:10.1038/nrn1787 (2005).

41 Sandberg, R. *et al.* Regional and strain-specific gene expression mapping in the adult mouse brain. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 11038-11043 (2000).