ALLEN INSTITUTE for BRAIN SCIENCE

Allen Mouse Brain Connectivity Atlas

TECHNICAL WHITE PAPER: TRANSGENIC CHARACTERIZATION

OVERVIEW

Transgenic mice are invaluable tools for the study of neural circuits and brain function. They enable the labeling of selective neuronal or non-neuronal populations, the modulation of gene expression in particular cell classes, and the manipulation of activity in specific cell populations. In 2007, the Allen Institute began to create a variety of transgenic mouse lines capable of expressing fluorescent probes, biological sensors, or functional proteins in a range of cell types throughout the brain. Based primarily on the Cre/loxP recombinase system, our efforts involve the creation of both reporter/responder mice that drive target genes in a Cre-dependent manner and Cre/CreERT2-driver mouse lines that express the recombinase from a transgenic or endogenous promoter in selected cell populations. To increase the level and specificity of reporter/responder gene expression, some recently generated mouse lines also employ elements of the tetracycline-inducible expression system, as well as recombinases other than Cre/CreERT2. In parallel with transgenic mouse generation, efforts have been made to qualitatively characterize gene expression of both driver and driver-dependent transgenes in lines generated by the Allen Institute or for lines created by other researchers but used in studies at the Allen Institute. Using standardized data collection protocols for histological profiling of transgene expression, this characterization data set is intended to provide a brain-wide profile for gene expression per mouse line, and thus assist researchers in choosing the appropriate transgenic tools for their studies of different regions and/or cell classes in the brain.

Currently, characterization data are available for approximately 25 Cre-dependent reporter lines generated at the Allen Institute and more than 200 Cre/CreERT2 or other driver lines, generated in-house or by other investigators using various expression strategies. Transgene expression in the reporter lines has been assessed by colorimetric *in situ* hybridization (CISH), double fluorescence *in situ* hybridization (DFISH) or native fluorescence of XFP (generic term for fluorescent proteins of different colors) in mice that were bred to also express the appropriate recombinase. The recombinase expression in the driver lines has been similarly evaluated by CISH, DFISH, native XFP fluorescence or immunohistochemical (IHC) labeling of reporter genes in mice that were bred to also express a recombinase-dependent fluorescent reporter.

The transgenic mouse lines generated by the Allen Institute are deposited to common repositories, such as The Jackson Laboratory, for distribution to the scientific community. Wherever available, the stock numbers are provided along with the characterization data. Currently, the mice that have been characterized are either on a mixed genetic background, mostly 129/B6, or on a congenic C57BL/6J background. For the lines that will be actively maintained, The Jackson Laboratory will backcross them into a C57BL/6J congenic background.

CHARACTERIZATION WORKFLOW

Using the systems developed for the generation of the Allen Mouse Brain Atlas, a characterization workflow was established to profile gene expression patterns throughout the entire mouse brain. The characterization

workflow has several advantages: (1) data are generated throughout the entire brain in a systematic way, (2) high resolution images are obtained from automated scanners, and (3) data registered to a reference atlas associated with the Allen Mouse Brain Atlas allows further analysis through integrated informatics. Gene expression in both Cre-dependent reporter and Cre driver transgenic lines is routinely characterized using this workflow as follows. The expression of reporter lines is assessed by crossing the reporter with various Cre driver lines, such as Emx1-Cre, Chat-Cre, Pcp2-Cre, etc., and analyzing adult offspring that are double transgenic. The Cre driver lines are then characterized by crossing the driver with a robust Cre-dependent reporter line, such as Ai14, and analyzing adult, as well as developing, offspring that carry both transgenes.

In addition to Cre driver lines that were made at the Allen Institute, we have also endeavored to characterize Cre lines obtained from external sources, including The Jackson Laboratory, NIH, GENSAT, and individual investigators. The lines from NIH include those that were generated as part of the NIH Neuroscience Blueprint Cre Driver Network. These lines are suitable for tissue- and time-specific perturbation of gene function in the nervous system. A number of Cre driver lines have been created and characterized by the Cre Driver Network and more than 90 of those are currently characterized. Characterization of these Cre lines is made possible through the NIH Blueprint contract R01DA028298. This characterization data is complementary to the NIH Neuroscience Blueprint Cre Driver Network data (http://www.credrivermice.org). In addition, approximately 30 BAC-Cre driver lines from GENSAT have also been characterized, complementing the GENSAT data (http://www.gensat.org/). Finally, we have characterized Cre driver lines generated by other individual investigators, which were obtained through The Jackson Laboratory, MMRRC or from the investigators directly.

Expression data from all Cre driver lines has been used to guide efforts toward constructing the Allen Mouse Brain Connectivity Atlas (the Atlas), one major aspect of which relies upon Cre expression in genetically-defined cell populations. All Cre driver characterization data is publicly available within the Atlas under the Transgenic Characterization tab.

Currently, our characterization workflow generates the following types of data: (1) ISH of Cre or the reporter gene to determine their mRNA expression throughout the brain. Because ISH signal is concentrated in cell bodies, the data provide single-cell resolution of the expression pattern. (2) DFISH of the reporter gene combined with a gene of interest (GOI), which can be the endogenous targeted gene or other cell-type marker genes. This kind of analysis can yield a variety of information depending on the probe pairs chosen. For example, using probes to the reporter and the endogenous gene, one can determine the extent to which reporter expression recapitulates that of the endogenous gene, as well as whether the reporter shows unexpected or ectopic expression. Alternatively, using probes to the reporter and various cell-type marker genes, one might also identify a particular cell class in which reporter expression is occurring. (3) XFP native fluorescence (*e.g.*, tdTomato or ZsGreen1) shows expression of the reporter gene at the protein level throughout the brain. If strong enough, the native fluorescence can reveal dendritic morphology and axonal projections of the labeled neurons. (4) IHC of the reporter gene (*e.g.*, EYFP) also shows expression of the reporter gene at the protein level throughout the brain, and if strong enough, can reveal dendritic morphology and axonal projections of the labeled neurons.

See the *Resources* white paper under the Documentation tab in the Allen Mouse Brain Connectivity Atlas for a complete list of all the transgenic lines (reporters or Cre drivers) that have been characterized.

CRE-DEPENDENT REPORTER LINES

To create a collection of mice that express fluorescent probes and functional tools at high levels in selected cell populations, cell-type specific Cre/CreERT2-driver lines have been combined with universally expressing Cre-dependent responder lines. For our initial work in reporter line generation, Cre-dependent transgenes have been targeted to the *Gt (ROSA) 26Sor* locus (Zambrowicz *et al*., 1997). Demonstrated to support nearly ubiquitous expression throughout the body, the Rosa26 locus has also been shown to be a permissive locus, in which strong exogenous promoters can direct high levels of transgene expression (Zong *et al*., 2005). Indeed, by targeting the Rosa locus with the reporter vectors depicted in **Figure 1**, a collection of mouse lines that

express, in a Cre-dependent manner, various proteins at high levels and in a wide range of cell types within the brain have been created. **Table 1** summarizes the Cre-dependent reporter lines for which we currently have characterization data.

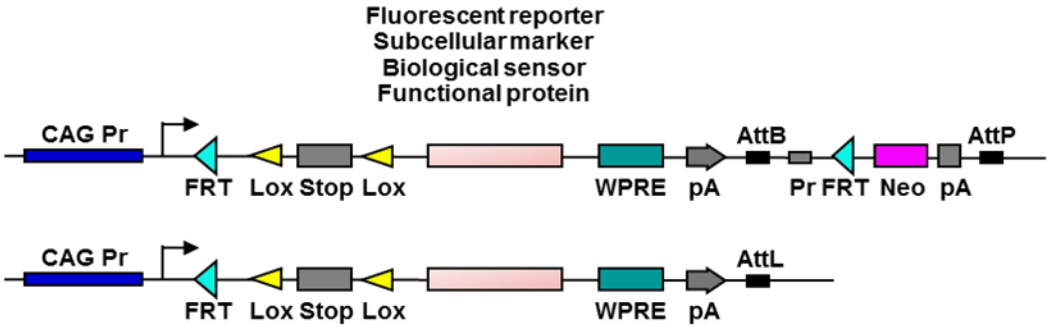


Figure 1. General targeting construct design for Cre-dependent reporter mice.
Genes are expressed from the CAG promoter following Cre-mediated deletion of the floxed stop cassette. The upper vector is the construct design transfected into embryonic stem (ES) cells and represents the actual insertion into the Rosa26 locus. The lower vector indicates the transgene configuration in the Rosa locus following removal of the selection cassette by PhiC31-mediated recombination of AttB and AttP sites.

Major elements of the Rosa26-targeting vectors include:

- CAG promoter: Strong and ubiquitously expressed CMV-chicken beta actin promoter.
- Lox-Stop-Lox (LSL) cassette: Prevents expression of the transgene in the absence of Cre-mediated recombination. In the presence of Cre recombinase, the LoxP sites recombine, and the stop cassette is deleted.
- Gene of interest: Includes fluorescent reporters, subcellular markers, biological sensors, and optogenetic tools. Lines with characterization data on the website are listed in **Table 1**.
- WPRE sequence: The woodchuck hepatitis virus post-transcriptional regulatory element is used in all but the Ai2 line to enhance mRNA transcript stability.
- Pair of FRT sites: Included to allow for a Flp-mediated replacement strategy.
- AttB/AttP sites: Recognition sites for the PhC31 recombinase. By flanking the PGK-Neomycin selection cassette with AttB and AttP sites, the cassette can be deleted *in vitro* or *in vivo* by PhiC31 recombinase (Raymond *et al*., 2007).

Double Reporter Lines

By making reporter gene expression dependent on the activity of Cre recombinase, it is possible to limit reporter expression to particular cell populations. In some circumstances, however, it is advantageous to restrict transgene expression to an even smaller group of cells than that defined by a single Cre-driver line. To achieve finer control over reporter expression, Rosa26-targeted reporter lines in which transgene expression is regulated by the activity of two independent recombinases have been created. Targeting vectors for these reporter lines contain all the elements previously used (**Figure 1**), but carry an additional recombinase-dependent stop cassette between the CAG promoter and the transgene. Vector designs for two lines that have been submitted to The Jackson Laboratory are shown in **Figure 2**. In the Ai65 line, expression of tdTomato requires the activity of both Cre and Flp recombinases, whereas in Ai66, both Cre and Dre recombinases are needed (Anastassiadis *et al*., 2009).

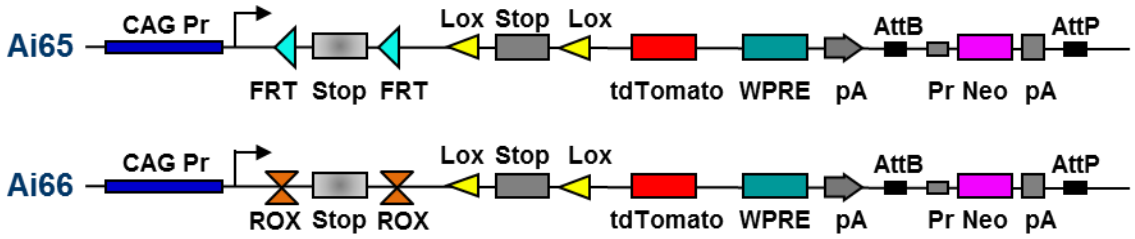


Figure 2. **Targeting construct design utilizing a double reporter strategy.**
The Rosa26 locus is targeted by constructs in which the tdTomato reporter gene is regulated by two different site specific recombinases. The Ai65 vector incorporates both the Cre/LoxP and the FLP/FRT recombinase systems, whereas Ai66 incorporates the Cre/LoxP and Dre/ROX systems.

Reporter Lines Based in the TIGRE Genomic Locus

Although many of our Rosa26-based reporter lines express functional levels of genetically encoded fluorescent probes and tools, we sought to develop an alternative expression strategy that would offer the potential for both increased specificity and enhanced levels of transgene expression. To do so, a docking site within the mouse TIGRE genomic locus on chromosome 9, which has been shown to enable strong tTA-inducible expression of transgenes in most tissues (Zeng *et al*., 2008), has been created. Using recombinase-mediated cassette exchange (RMCE) (Turan *et al*., 2011) in a correctly targeted ES cell clone, a variety of genetically encoded markers, sensors and effectors, whose expression is dually regulated by both Cre and tTA, have been introduced into the locus. The composition of the TIGRE transgene allele in the Ai62 line, which contains the sequence encoding the tdTomato fluorescent protein, is depicted in **Figure 3**. Components of the allele in other TIGRE reporter lines are identical to that shown except for the transgene expressed. TIGRE-based reporter lines with characterized expression data are listed in **Table 1**.

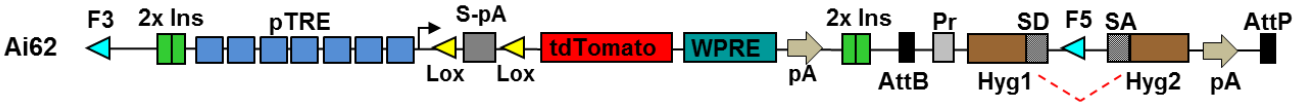


Figure 3. **The TIGRE genomic allele in the Ai62 tdTomato reporter line following RMCE into a pretargeted docking site.**
Transgenes are expressed from the tetO promoter, pTRE, following Cre-mediated deletion of the floxed stop cassette.

Major elements of TIGRE reporter alleles include:

- FRT3 and FRT5 sites: Included to allow for RMCE in correctly targeted ES cell clones. By incorporating this strategy, a variety of functional elements can readily be introduced into the TIGRE locus without the need for repeated targeting by homologous recombination.
- Two pair of tandem repeats of the 1.2 kb chicken β-globin HS4 insulator elements: Insulators shield the expression cassette in between them from silencing chromatin modification.
- pTRE: the tTA/rtTA-regulated tetO promoter unit is comprised of seven repeated binding sites and a minimal CMV promoter.
- Lox-Stop-Lox (LSL) cassette: Prevents expression of the transgene in the absence of Cre-mediated recombination.
- Gene of interest: Includes fluorescent reporters, biological sensors and optogenetic effectors. Lines with characterization data on the website are listed in **Table 1**.
- WPRE sequence: The woodchuck hepatitis virus post-transcriptional regulatory element is included to enhance mRNA transcript stability.
- AttB/AttP sites: Recognition sites for the PhC31 recombinase. By flanking the PGK-Hygromycin selection cassette with AttB and AttP sites, the cassette can be deleted *in vitro* or *in vivo* by PhiC31 recombinase.

- Split Hygromycin gene (Hyg1 and Hyg2 domains) and linked mRNA splice donor (SD) and splice acceptor (SA) sequences: The initial docking site contains the SA-Hyg2 unit while the incoming vector for the RMCE contains the Hyg1-SD unit. The reconstituted Hygromycin gene allows for drug selection of correct RMCE reactions.

Table 1. Reporter lines developed by the Allen Institute for Brain Science.

Reporter Line	Expressed Gene	Description of Functionality	Promoter and Knock-in Locus	Regulation	Reference
Ai2	EYFP	Fluorescent marker	pCAG in Rosa26	Cre	Madisen et al., 2010
Ai3	EYFP	Fluorescent marker	pCAG in Rosa26	Cre	Madisen et al., 2010
Ai6	ZsGreen1	Fluorescent marker	pCAG in Rosa26	Cre	Madisen et al., 2010
Ai9	tdTomato	Fluorescent marker	pCAG in Rosa26	Cre	Madisen et al., 2010
Ai14	tdTomato	Fluorescent marker	pCAG in Rosa26	Cre	Madisen et al., 2010
Ai27	ChR2 (H134R)-tdTomato	Channelrhodopsin mutant fused to tdTomato	pCAG in Rosa26	Cre	Madisen et al., 2012
Ai31	Syp-EmGFP	Synaptophysin fused to EmGFP	pCAG in Rosa26	Cre	Nakata et al., 1998
Ai32	ChR2 (H134R)-EYFP	Channelrhodopsin mutant fused to EYFP	pCAG in Rosa26	Cre	Madisen et al., 2012
Ai34	Syp-tdTomato	Synaptophysin fused to tdTomato	pCAG in Rosa26	Cre	Nakata et al., 1998
Ai35	Arch-EGFP-ER2	Archaerhodopsin fused to EGFP	pCAG in Rosa26	Cre	Madisen et al., 2012
Ai38	GCaMP3	Calcium indicator	pCAG in Rosa26	Cre	Zariwala et al., 2012
Ai39	eNpHR3.0-EYFP	Halorhodopsin fused to EYFP	pCAG in Rosa26	Cre	Madisen et al., 2012
Ai40	ArchT-EGFP	Archaerhodopsin fused to EGFP	pCAG in Rosa26	Cre	Han et al., 2011
Ai47	EmGFP-T2A-TagGFP2-P2A-hrGFP	Fluorescent markers	pCAG in Rosa26	Cre	Steinecke et al., 2017
Ai57	Jaws-GFP-ER2	Cruxhalorhodopsin fused to EGFP	pCAG in Rosa26	Cre	Madisen et al., 2015
Ai62	tdTomato	Fluorescent marker	TRE in TIGRE	Cre and tTA	Madisen et al., 2015
Ai63	tdTomato	Fluorescent marker	TRE in TIGRE	tTA	
Ai65	tdTomato	Fluorescent marker	pCAG in Rosa26	Cre and Flp	Madisen et al., 2015
Ai65F	tdTomato	Fluorescent marker	pCAG in Rosa26	Flp	
Ai66	tdTomato	Fluorescent marker	pCAG in Rosa26	Cre and Dre	Madisen et al., 2015
Ai75	nls-tdTomato	Nuclear localized fluorescent marker	pCAG in Rosa26	Cre	
Ai78	VSFPB	Voltage indicator VSFP Butterfly 1.2	TRE in TIGRE	Cre and tTA	Madisen et al., 2015

Ai79	Jaws-GFP-ER2	Cruxhalorhodopsin fused to EGFP	TRE in TIGRE	Cre and tTA	Madisen et al., 2015
Ai82	EGFP	Fluorescent marker	TRE in TIGRE	Cre and tTA	Madisen et al., 2015
Ai85	iGluSnFR	Glutamate indicator	TRE in TIGRE	Cre and tTA	Madisen et al., 2015
Ai86	ArcLight	Voltage indicator	TRE in TIGRE	Cre and tTA	Jin et al., 2012
Ai87	iGluSnFR	Glutamate indicator	pCAG in Rosa26	Cre	Madisen et al., 2015
Ai90	Chronos	Channelrhodopsin Chronos fused to EGFP	TRE in TIGRE	Cre and tTA	
Ai92	YCX2.60	Calcium indicator	TRE in TIGRE	Cre and tTA	Madisen et al., 2015
Ai93	GCaMP6f	Calcium indicator	TRE in TIGRE	Cre and tTA	Madisen et al., 2015
Ai94	GCaMP6s	Calcium indicator	TRE in TIGRE	Cre and tTA	Madisen et al., 2015
Ai95	GCaMP6f	Calcium indicator	pCAG in Rosa26	Cre	Madisen et al., 2015
Ai96	GCaMP6s	Calcium indicator	pCAG in Rosa26	Cre	Madisen et al., 2015
Ai136	ReaChR-EYFP	Red-shifted channelrhodopsin variant fused to EYFP	TRE in TIGRE	Cre and tTA	unpublished
Ai148	GCaMP6f	Calcium indicator	TRE2 and pCAG in TIGRE	Cre	unpublished
Snap25-T2A-GCaMP6s	GCaMP6s	Calcium indicator	Snap25 endogenous promoter and locus	Cre and endogenous promoter	Madisen et al., 2015
Snap25-LSL-F2A-GFP	EGFP	Fluorescent marker	Snap25 endogenous promoter and locus	Cre and endogenous promoter	Madisen et al., 2015
Chrm2-tdT	tdTomato	Fluorescent marker	Chrm2 endogenous promoter and locus	Endogenous promoter	unpublished

The targeting constructs described above were transfected into G4 ES cells, a 129/B6 F1 hybrid ES cell line (George *et al*., 2007). Transfection into G4 cells over the inbred 129 or B6 ES cell lines was selected because the G4 line has been reported to give especially robust germline transmission along with a high percentage of chimeras. Our experience has confirmed this.

Expression and/or functional characterization of the Cre-dependent reporter lines can be found at: Ai2-Ai14 (Madisen *et al*., 2010), Ai27, Ai32, Ai35, Ai39 (Madisen *et al*., 2012), Ai38 (Zariwala *et al*., 2012) and Ai62, Ai65, Ai66, Ai78, Ai82, Ai87, Ai95 and Ai96 (Madisen *et al*., 2015).

DRIVER LINES

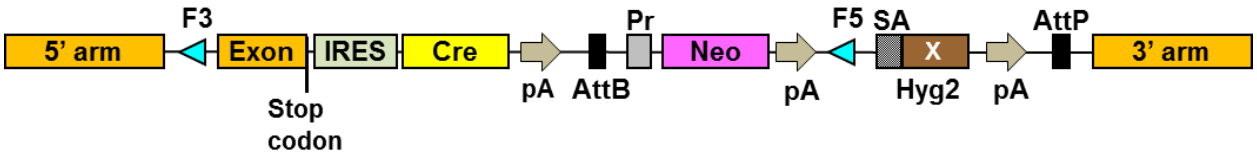
As part of our effort to create transgenic lines capable of expressing markers and genetic tools in a range of specific cell types throughout the brain, Cre, CreER, CreERT2, dCre, tTA2, FlpO, and dgFlpO driver lines using either BAC or knock-in approaches have been generated.

Generation of Knock-in Driver Lines

To establish cell-type or cell-population specific driver lines, a knock-in approach to target a given recombinase or tTA2 to specific genetic loci has been taken. Targets chosen for knock-in comprise a diverse set of genes,

including: genes that could serve as interneuron markers; genes enriched in particular cell-types, cortical layers or structures; genes whose expression mark pre- or post-synaptic termini; and genes with otherwise intriguing expression patterns (**Table 2**). Knock-in vectors have been designed to target primarily the endogenous gene's translational stop site, where either an IRES or a 2A "self-cleaving" sequence upstream of the driver gene (*i.e.*, last exon of gene-2A-driver) has been utilized to minimize the disruption of endogenous gene expression. In a few other cases, vectors target either the endogenous gene's translational start site with a driver or driver-2A cassette, or they target an intron with a splice acceptor-2A-driver containing cassette. In most instances, targeting vectors introduce the driver gene sequence directly into the genomic locus, although for a few lines, a docking site to a genomic locus has been targeted, followed by use of RMCE in a correctly targeted ES cell clone to introduce the driver gene of choice. Examples of genomic alleles generated through knock-in targeting are depicted in **Figure 4**.

A



B

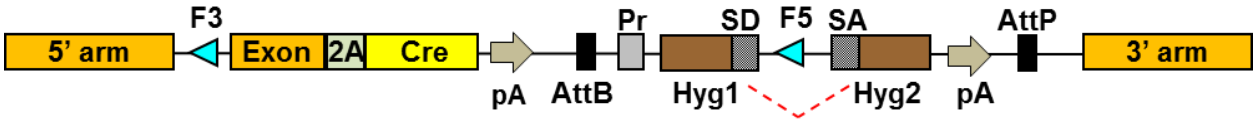


Figure 4. Representative configurations of targeted genomic loci in Cre driver lines.
An IRES-Cre driver is directly targeted to a genomic locus in **(A)** and RMCE has been used to introduce a 2A-Cre driver into a pretargeted docking site in **(B)**.

Major elements of the knock-in driver vectors include:

- Homology arms: The 5'- and 3'- homology arms contain 2-6 kb of mouse genomic sequence isolated from a BAC clone (RP23 or KOMP libraries) or chemically synthesized, based on publically available sequence of the C57BL/6 genome.
- FRT3 and FRT5 sites: Included to allow for RMCE in correctly targeted ES-cell clones. By incorporating this strategy, a variety of functional elements can readily be introduced into a specific locus without the need for repeated targeting by homologous recombination.
- IRES or 2A sequences: Allow for the production of driver protein from a bicistronic mRNA. The IRES (internal ribosomal entry site) element is introduced after the targeted gene's endogenous stop codon, and it recruits a new ribosome to the transcript. The 2A"self-cleaving" peptide acts as translation-interruption/reinitiation sequence, and it is introduced to replace the endogenous stop codon in frame with the upstream coding sequence.
- Driver gene: Generally Cre, or CreERT2, however some lines are designed to express versions of destabilized Cre (DHFR-Cre or DHFR-GFP-Cre), which require the presence of trimethoprim (TMP) for protein stability and optimal activity (Sando *et al*., 2013), tTA2 (Urlinger *et al*., 2000), FlpO (Raymond *et al*., 2007) or destabilized FlpO (DHFR-GFP-FlpO).
- AttB/AttP sites: Recognition sites for the PhC31 recombinase. By flanking the selectable marker cassette with AttB and AttP sites, the cassette can be deleted *in vitro* or *in vivo* by PhiC31 recombinase.
- Promoter-Neomycin/promoter hygromycin-pA: Selectable marker cassette for the gene-targeting transfection or the RMCE into a previously targeted ES cell clone.

- Split Hygromycin gene (Hyg1 and Hyg2 domains) and linked mRNA splice donor (SD) and splice acceptor (SA) sequences: The initial docking site contains the SA-Hyg2 unit while the incoming vector for the RMCE contains the Hyg1-SD unit. The reconstituted Hygromycin gene allows for drug selection of correct RMCE reactions.

Table 2. Allen Institute for Brain Science knock-in driver lines.

Gene Name	Abbreviation	Driver	Insertion site
Adenylate cyclase activating polypeptide 1	Adcyap1	Cre	Stop codon
Arginine vasopressin	Avp	Cre	Stop codon
Calbindin 1	Calb1	dCre	Stop codon
Calbindin 1	Calb1	Cre	Stop codon
CART prepropeptide	Cart	Cre	Stop codon
Connective tissue growth factor	Ctgf	dCre	Stop codon
Erythroblastic leukemia viral oncogene homolog 4 (avian)	Erbb4	CreERT2	Stop codon
Estrogen receptor 2	Esr2	Cre	Stop codon
Fez family zinc finger 1	Fezf1	dCre	Stop codon
G protein subunit beta 4	Gnb4	Cre	Stop codon
G protein subunit beta 4	Gnb4	CreERT2	Stop codon
5-hydroxytryptamine receptor 1A	Htr1a	Cre	Stop codon
Natriuretic peptide receptor 3	Npr3	Cre	Stop codon
Netrin G2	Ntng2	Cre	Stop codon
Neuropeptide Y	Npy	FlpO	Stop codon
Neurexophilin 4	Nxph4	CreERT2	Stop codon
Neuron-derived neurotrophic factor	Ndnf	dCre	Stop codon
Oxytocin receptor	Oxtr	Cre	Stop codon
Parvalbumin	Pvalb	Cre	Stop codon
Parvalbumin	Pvalb	CreERT2	Stop codon
Parvalbumin	Pvalb	dCre	Stop codon
Parvalbumin	Pvalb	Dre	Stop codon
Parvalbumin	Pvalb	FlpO	Stop codon
Parvalbumin	Pvalb	tTA2	Stop codon
Plexin-D1 precursor	Plxnd1	dgFlpO	Stop codon

Prodynorphin	Pdyn	CreERT2	Stop codon
Preproenkephalin	Penk	Cre	Stop codon
Preproenkephalin	Penk	CreERT2	Stop codon
RAR-related orphan receptor beta	Rorb	Cre	Stop codon
RAR-related orphan receptor beta	Rorb	tTA2	Stop codon
RAR-related orphan receptor beta	Rorb	FlpO	Stop codon
Ras protein-specific guanine nucleotide releasing factor 2	Rasgrf2	dCre	Stop codon
Ras protein-specific guanine nucleotide releasing factor 2	Rasgrf2	dgFlpO	Stop codon
Solute carrier family 17 member 6	Slc17a6 (Vglut2)	FlpO	Stop codon
Sodium-dependent inorganic phosphate cotransporter, member 7	Slc17a7 (Vglut1)	Cre	Stop codon
Solute carrier family 17 member 8	Slc17a8 (Vglut3)	Cre	Stop codon
Solute carrier family 32 member 1	Slc32a1 (VGAT)	FlpO	Stop codon
Somatostatin	Sst	Cre	ATG start
Synaptosomal-associated protein 25	Snap25	Cre	Stop codon
Tachykinin 1	Tac1	Cre	Stop codon
Tachykinin 2	Tac2	Cre	Stop codon
Tachykinin receptor 1	Tacr1	Cre	Stop codon
Tribbles homolog2 (Drosophila)	Trib2	CreERT2	Intron 1
Vasoactive intestinal polypeptide receptor 2	Vipr2	Cre	Stop codon

Fusion proteins are abbreviated as dCre (DHFR-Cre).

Generation of BAC Cre Drivers

To produce BAC transgenic mice that express Cre in a regionally and spatially-restricted manner, a set of genes with expression patterns that are restricted to specific cell populations (cortical layers or inhibitory neuron subtypes) with as much global specificity for the neocortex as possible were selected. Candidate genes (**Table 3**) were reviewed for any information available (including existing transgenic animals, literature, genomic structure and position, splice variants, and start site) and other databases (BGEM, GENSAT, and Unigene) were examined to obtain confirming expression data when available.

Plasmids obtained containing Cre and inducible CreERT2 (kindly provided by Professor Pierre Chambon) were targeted with a selectable marker flanked by FRT sites (FRT-Neo-FRT). These initial plasmids were used for subsequent targeting to BACs that contained genes of interest. The Cre FRT-Neo-FRT portions of the plasmids were PCR amplified with flanking oligonucleotides designed to have 50 base pair-long homology arms with the BAC of interest. RedET recombination was used to insert the sequence-verified PCR product into the BAC locus (Zhang *et al*., 1998). The Neomycin selection cassette was subsequently removed from the BAC by FLP-mediated recombination. The wild type loxP site present in the RP23 BAC backbone was removed before pronuclear injection by restriction enzyme digestion and fragment purification. Plasmid construction was produced in collaboration with Gene Bridges GmbH (Heidelberg, Germany).

Preparation of large quantities of linearized BAC DNA for pronuclear injection made use of the commercially available Nucleobond BAC100 (Clontech, USA). BACs were linearized and purified over a Sepharose CL4B column. Fractions were analyzed by pulse field gel electrophoresis to identify the sample with the highest concentration of linearized BAC DNA and lowest concentration of vector DNA. Linearized BAC DNA was injected at a concentration of 1 µg/ml into C57BL/6 x B6C3F1 zygotes.

Table 3. BAC transgenic Cre driver lines.

Driver	BAC	Gene	Neocortical Specificity
A9300038C07RIK-Cre	RP23-476E22	RIKEN cDNA A9300038C07 gene	Layer 1
Wfs1-CreERT2	RP23-405O19	Wolfram syndrome 1 homolog (human)	Layer 2
Cart-Cre	RP23-379C17	CART prepropeptide	Layer 3
Scnn1a-Cre	RP23-405B24	Sodium channel, nonvoltage-gated, type I, alpha	Layer 4
Cyp39a1-Cre	RP23-370 M9	Cytochrome P450, family 39, subfamily a, polypeptide 1	Layer 4
Hsd11b1-Cre	RP23-451H2	Hydroxysteroid 11-beta dehydrogenase 1	Layer 5
Ctgf-Cre	RP23-257E12	Connective tissue growth factor 1	Layer 6
Mybpc1-Cre	RP23-116K19	Myosin binding protein C (RIKEN cDNA 8030451F13) gene	Scattered

DATA COLLECTION METHODS FOR HISTOLOGICAL CHARACTERIZATION

Tamoxifen Induction

Mice generated with tamoxifen inducible CreERT2 constructs were housed under normal conditions for 1-2 months before being treated with tamoxifen. Tamoxifen solution was prepared by mixing tamoxifen powder in corn oil at a final concentration of 50 mg/ml. The solution was placed on a heated stir plate set to 40°C and stirred until tamoxifen was completely dissolved. Mice at age P28 or above were administered approximately 100 µl tamoxifen solution (200 mg/kg) via oral gavage once a day for 5 days. Mice at age P4 or P14 were given a single dose of 40 µl tamoxifen solution at age P2 and P10, respectively. Animals were monitored for adverse effects, and if these become apparent, treatment was stopped. One week after treatment ended, animals were processed as described below for ISH, DFISH, or localization of XFP.

Trimethoprim Induction

Mice generated with constructs containing destabilized Cre drivers (dCre) were treated with trimethoprim to stabilize the driver protein prior to analysis. Animals were housed under normal conditions for 1-2 months before treatment. Trimethoprim solution was prepared by mixing trimethoprim powder in DMSO liquid to a concentration of 100 mg/ml. Then, trimethoprim was diluted to 20 mg/ml by adding a 2% methylcellulose solution. The solution was then ready to administer to mice at a volume of 0.015 ml/g of body weight. Mice at age P28 or above were administered 20 mg/ml via oral gavage once a day for 3 days, and brains were collected for analysis after 7 or more days post-treatment.

Colorimetric *In Situ* Hybridization
The systems developed for processing tissue, RNA *in situ* hybridization (ISH), Nissl staining, image acquisition, and data processing for the generation of the Allen Mouse Brain Atlas were used. The framework, workflow and equipment are described in Lein *et al*., 2007 as well as in the *Data Production Processes* white paper in the Documentation section of the Allen Mouse Brain Atlas.

Double Fluorescence *In Situ* Hybridization (DFISH)
Double fluorescence *in situ* hybridization (DFISH) is based on the colorimetric ISH protocol and was performed as previously described (Thompson *et al*., 2008). A VS110/VS120 fluorescence scanning system (Olympus Center Valley, PA) equipped with a 10x objective (N.A. 0.4) is utilized for image acquisition, instead of the original Image Capture Systems (ICS).

Tape Transfer Sectioning
Tape transfer sectioning was done on a Leica 3050 S Cryostat equipped with an Instrumedics Tape Transfer System that includes a UV light polymerization chamber and warming pad. A roll of Instrumedics tape flags (Instrumedics cat. #TW) was cut into individual flags. Slides were coated with 50 µL of Solution B adhesive (Instrumedics cat. #475272) and allowed to air dry. Prior to taking a section, slides were positioned on the warming pad. The tacky surface of the tape flag was positioned over the blockface of the tissue. Sections were cut and placed on the slide on the warming pad. A brush was used to apply pressure to the tape flag so that it was thoroughly laminated to the slide. After all tape flags for a slide were laminated, the slide was placed into the UV light polymerization unit. Following UV light treatment, the tape flags were removed with forceps, leaving the sections attached to the slide.

Localization of XFP
Mice were anesthetized with 5% isoflurane and intracardially perfused with 10 ml of saline (0.9% NaCl) followed by 50 ml of freshly prepared 4% paraformaldehyde (PFA) at a flow rate of 9 ml/min. Brains were rapidly dissected and post-fixed in 4% PFA at room temperature for 3-6 hours and overnight at 4°C. Brains were cryoprotected in 30% (w/v) sucrose in PBS at 4°C overnight or until they sunk to the bottom of the tube. The brains were embedded in optimal cutting temperature compound (OCT) and 25 µm cryostat sections were cut using tape transfer methods. Slides were air dried for 30 min, and then stored at -80°C. For imaging, sections were removed from -80°C and allowed to dry at room temperature. Sections were DAPI stained and coverslipped using Hydromatrix supplemented with 5.0% Dabco antifade (Sigma cat. #290734). The slides were imaged as described for DFISH above.

Immunohistochemistry
Tissue specimens were prepared as described in the Localization of XFP section. For single-label immunofluorescence, sections were removed from -80°C and allowed to dry at room temperature. Slides were incubated for 30 min at room temperature in PBS containing 0.3% Triton X-100 and 5% normal goat serum (NGS). Polyclonal antibody against GFP (Abcam cat. #ab6556) was diluted 1:1000 in PBS containing 0.3% Triton X-100 and 5% normal goat serum. Sections were coverslipped with parafilm and incubated overnight at 4°C in a humid chamber. The sections were rinsed for 30 min in PBS containing 1% NGS, incubated in goat anti-rabbit IgG-Alexa Fluor 488; (Invitrogen, 1:400 dilution) for 2 hr at room temperature, rinsed for 10 min in PBS containing 1% NGS, and rinsed for 30 min in PBS. Sections were DAPI stained and coverslipped using Hydromatrix supplemented with 5.0% Dabco antifade (Sigma cat. #290734). The slides were imaged as described for DFISH.

Data Processing
Once images were acquired, the Informatics Data Pipeline (IDP) managed image preprocessing, image quality control (QC), ISH expression detection and measurement, annotation quality control (QC) and public display of information via the Web application. The IDP has been described in detail previously (Dang *et al*., 2007), and has been extended to process images for this project. In the preprocessing step, bright field images are white-balanced and a tissue detection algorithm assigns bounding boxes to individual tissue sections, which are manually assessed and adjusted when necessary. For fluorescent images, the intensity of each color channel

has been transformed (window/level) to enable better visualization of the information content. This is followed by tissue detection to generate bounding boxes.

For ISH images, an adaptive image processing algorithm was applied to detect and quantify gene expression. This processing module creates an expression mask of detected “expressor” objects which is pseudo-color coded for display to provide a gestalt view of expression levels across the tissue. Each image set is registered to an age-matched 3D reference model. Transforms (12 parameters affine) from the registration process are stored in the database and used in the Web application to support the image synchronization function in the Image Viewer. For adult data, the 3D reference model is built upon the coronal plates of the reference atlas associated with the Allen Mouse Brain Atlas, enabling the creation of gene expression graphs in the experiment details page for large anatomical areas.

Quality Control

Quality control procedures are implemented for data acquisition and analysis, similar to protocols observed for other Allen Brain Atlas data sets. In general, all image series are inspected for artifacts which may reduce consistency, analyzability, or completeness of the data. Artifacts such as missing tissue, misaligned section placement, poor staining quality, edge cutoff, sections flipped, and low signal strength may cause image series to be excluded.

REFERENCES

Anastassiadis K, Fu J, Patsch C, Shengbiao H, Weidlich S, Duerschke K, Buchholz F, Edenhofer F, Stewart AF (2009) Dre recombinase, like Cre, is a highly efficient site-specific recombinase in *E. coli*, mammalian cells and mice. *Dis Models & Mech* 2:508-515.

George SH, Gertsenstein M, Vintersten K, Korets-Smith E, Murphy J, Stevens ME, Haigh JJ, Nagy A (2007) Developmental and adult phenotyping directly from mutant embryonic stem cells. *Proc Natl Acad Sci U S A* 104:4455-4460.

Han X, Chow BY, Zhou H, Klapoetke NC, Chuong A, Rajimehr R, Yang A, Baratta MV, Winkle J, Desimone R, Boyden ES (2011) A high-light sensitivity optical neural silencer: development and application to optogenetic control of non-human primate cortex. *Front Syst Neurosci* 13;5:18.

Jin L, Han Z, Platisa J, Wooltorton JR, Cohen LB, Pieribone VA (2012) Single action potentials and subthreshold electrical events imaged in neurons with a fluorescent protein voltage probe. *Neuron* 75:779-785.

Lein ES *et al*. (2007) Genome-wide atlas of gene expression in the adult mouse brain. *Nature* 445:168-176.

Madisen L, Zwingman T, Sunkin S, Oh SW, Zariwala H, Gu H, Ng LL, Palmiter RD, Hawrylycz MJ, Jones AR, Lein ES, Zeng H (2010) A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nature Neuroscience* 13:133-140.

Madisen L, Mao T, Koch H, Zhuo J, Berenyi A, Fujisawa S, Hsu YW, Garcia AJ 3rd, Gu X, Zanella S, Kidney J, Gu H, Mao Y, Hooks BM, Boyden ES, Buzsaki G, Ramirez JM, Jones AR, Svoboda K, Han X, Turner EE, Zeng H (2012) A toolbox of Cre-dependent optogenetic transgenic mice for light-induced activation and silencing. *Nature Neuroscience* 15:793-802.

Madisen L, Garner A, Shimaoka D, Chuong A, Klapoetke N, Li L, van der Bourg A, Niino Y, Egolf L, Monetti C, Gu H, Mills M, Cheng A, Tasic B, Nguyen T, Sunkin S, Benucci A, Nagy A, Niyawaki A, Helmchen F, Empson RM, Knöpfel T, Boyden E, Reid C, Carandini M, Zeng H (2015) Transgenic mice for intersectional targeting of neural sensors and effectors with high specificity and performance. *Neuron* 85:942-958.

Nakata T, Terada S, Hirokawa N (1998) Visualization of the dynamics of synaptic vesicle and plasma membrane proteins in living axons. *J Cell Biol* 140:659–674.

Raymond CS, Soriano P (2007) High-efficiency FLP and PhiC31 site-specific recombination in mammalian cells. *PLoS ONE* 2:e162.

Sando R, Baumgaertel K, Pieraut S, Torabi-Rander N, Wandless TJ, Mayford M, Mximov A (2013) Inducible control of gene expression with destabilized Cre. *Nature Methods* 10:1085-1088.

Steinecke A, Hozhabri E, Tapanes S, Ishino Y, Zeng H, Kamasawa N, Taniguchi H (2017) Neocortical Chandelier Cells Developmentally Shape Axonal Arbors through Reorganization but Establish Subcellular Synapse Specificity without Refinement. *eNeuro* 12;4(3).

Thompson CL, Pathak SD, Jeromin A, Ng LL, MacPherson CR, Mortrud MT, Cusick A, Riley ZL, Sunkin SM, Bernard A, Puchalski RB, Gage FH, Jones AR, Bajic VB, Hawrylycz MJ, Lein ES (2008) Genomic anatomy of the hippocampus. *Neuron* 60:1010-1021.

Turan S, Galla M, Ernst E, Qiao J, Voelkel C, Schiedlmeier B, Zehe C, Bode J (2011) Recombinase-mediated cassette exchange (RMCE): traditional concepts and current challenges. *J Mol Biol* 407:193-221.

Urlinger S, Baron U, Thellmann M, Hasan MT, Bujard H, Hillen W (2000) Exploring the sequence space for tetracycline-dependent transcriptional activators: Novel mutations yield expanded range and sensitivity. *Proc Natl Acad Sci USA* 97:7963-7968.

Zambrowicz BP, Imamoto A, Fiering S, Herzenberg LA, Kerr WG, Soriano P (1997) Disruption of overlapping transcripts in the ROSA beta geo 26 gene trap strain leads to widespread expression of beta-galactosidase in mouse embryos and hematopoietic cells. *Proc Natl Acad Sci U S A* 94:3789-3794.

Zariwala HA, Borghuis BG, Hoogland TM, Madisen L, Tian L, DeZeeuw CI, Zeng H, Looger LL, Svoboda K, Chen T-W (2012) A Cre-dependent GCaMP3 reporter mouse for neuronal imaging in vivo. *J Neuroscience* 32:3131-3141.

Zeng H, Horie K, Madisen L, Pavlova MN, Gragerova G, Rohde AD, Schimpf B, Liang Y, Ojala E, Kramer F, Roth P, Slobodskaya O, Dolka I, Southon EA, Tessarollo L, Bornfeldt KE, Gragerov A, Pavlakis GN, Gaitanaris GA (2008) An inducible and reversible mouse genetic rescue system. *PLoS Genetics* 4(5):e1000069.

Zhang Y, Buchholz F, Muyrers JP, Stewart AF (1998) A new logic for DNA engineering using recombination in *E. coli*. *Nature* 20:123–128.

Zong H, Espinosa JS, Su HH, Muzumdar MD, Luo L (2005) Mosaic analysis with double markers in mice. *Cell* 121:479-492.