Co-opting Intracellular Proteins for Cell-Specific Gene Manipulation

A dissertation presented

by

Chung Yiu Jonathan Tang

to

The Division of Medical Sciences

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

In the subject of

Genetics

Harvard University

Cambridge, Massachusetts

November, 2014

© 2014 Chung Yiu Jonathan Tang

All Rights Reserved

Dissertation Advisor: Dr. Constance Cepko Chung Yiu Jonathan Tang

Co-opting Intracellular Proteins for Cell-Specific Gene Manipulation

Abstract

Studies of complex multicellular organisms would benefit from the ability to selectively manipulate the activities of any cell type of interest. Our ability to achieve this is currently limited by technology and available resources. Here, I explore the artificial use of intracellular proteins as signals for conferring cell specificity in gene manipulation.

The Green Fluorescent Protein (GFP) is a useful marker of gene expression and thousands of transgenic GFP reporter lines have been made to label different cell types, particularly in the mouse nervous system. However, the utility of transgenic GFP reporter lines is limited to labeling purposes. I exploited this resource for cell-specific gene manipulation by constructing synthetic systems that become biologically active upon interaction with GFP. Using GFP-binding nanobodies derived from Camelid antibodies, I co-opted GFP as a scaffold protein to bring together complementary split proteins that, when in a complex, can regulate processes such as transcription and DNA recombination. I demonstrate the utility of these systems for selectively manipulating GFP-expressing cells in the mouse nervous system and in zebrafish, for applications such as developmental perturbations, electrophysiology and optogenetic interrogation of neural circuits.

To reduce the complexity of GFP-dependent systems, I developed a binary system in which GFP binds to a destabilized nanobody and, in doing so, stabilizes expression level of an output protein fused to the nanobody. I show that this approach could be used

to construct fluorescent sensors and functional effectors that are only active in the presence of the intracellular antigen. This strategy can be extended to target a different intracellular protein, the HIV C-terminal domain (CTD), for intracellular antigen-inducible protein stabilization. The strategy used to generate CTD-inducible protein sensors may potentially be generalized to facilitate rapid design of protein-responsive sensors and effectors, based on elucidation of an amino acid code that can be applied across conserved protein scaffolds or antibody frameworks regardless of antigen identity.

Thus, my work expands the experimental paradigm for manipulation of specific cell types in multicellular organisms and provides tools and approaches for increasingly precise analysis of biological processes in transgenic and wildtype animals.

Table of Contents

Abstract iii

Table of Contents v

List of Figures and Tables viii

Acknowledgements xi

Dedication xiii

Chapter One: Introduction 1

A Reductionist approach to dissecting complex tissues 1

Defining cell types 1

Genetic tools facilitating the study of cell types 7

Green Fluorescent Protein 11

Limitation of existing GFP tools, and approach to cell-specific gene manipulation 16

Nanobody reagents for use with Green Fluorescent Protein 17

Notable features of nanobodies 20

Strategies for controlling biological activities with proteins – concepts related to small molecule regulated systems 20

Retina as a model system for testing engineered protein-responsive constructs 26

Summary of Thesis 28

References 29

Chapter Two: A Nanobody-Based System Using Fluorescent Proteins as Scaffolds for Cell-Specific Gene Manipulation.....32

Summary.....34

Introduction.....34

Results.....37

Discussion.....56

Experimental Procedures.....65

References.....68

Chapter Three: A GFP-Dependent Cre Recombinase for Retrofitting Transgenic GFP Lines.....73

Summary.....75

Introduction.....75

Results.....76

Discussion.....98

Material and Methods.....102

References.....105

Chapter Four: Detection and Manipulation of Antigen-Expressing Cells with Destabilized Binding Proteins.....107

Summary.....109

Introduction.....109

Results.....112

Discussion 134

Material and Methods 140

References 145

Chapter Five: Concluding Discussion 147

Fluorescent proteins as multifunctional switches, and protein-responsive systems 148

Natural function of fluorescent proteins and other molecular tools 151

Exploiting endogenous, intracellular products for cell-specific manipulations ... 153

Summary 155

References 157

Appendix I: Supplemental Information for Chapter 2 158

Appendix II: Rejected Cover Art 173

List of Figures and Tables

Chapter One: Introduction

Figure 1.1. Diversity of cell types in the typical mammalian retina.....3

Figure 1.2. Targeting cells by exploiting endogenous processes.....5

Figure 1.3. Binary systems and strategies commonly used to study genetically tractable model organisms. 8

Figure 1.4. The development of Green Fluorescent Protein as a tool.....12

Figure 1.5. Diverse transgenic GFP reporter lines have been generated and characterized for the mouse community. 14

Figure 1.6. Exploiting transgenic reporter lines as tools for cell-specific gene manipulation. 18

Figure 1.7. GFP-binding nanobodies. 21

Figure 1.8. Camelid antibodies and nanobodies. 22

Figure 1.9. Two possible strategies exploiting intracellular proteins as signals to control gene expression. 24

Figure 1.10. Organization of the retina and diverity of GFP lines labeling specific cell types. 27

Chapter Two: A Nanobody-Based System Using Fluorescent Proteins as Scaffolds for Cell-Specific Gene Manipulation

Figure 2.1. **.** In Vitro Screen Used to Identify Functional GBP Pairs for the GFP-Dependent Transcription System. 39

Figure 2.2. Characterization of the GFP-Dependent Transcription System. 41

Figure 2.3. T-DDOGs Are Highly Adjustable.....43

Figure 2.4. GFP Controls the Spatial Expression of Genes In vivo. 46

Figure 2.5. T-DDOGs Support Electrophysiological and Gene Perturbation Studies in the Central Nervous System. 50

Figure 2.6. Retrofitting a Transgenic GFP Mouse Line for GFP-Dependent Manipulation of Gene Expression and Neural Circuit Activities.....53

Figure 2.7. GFP-Dependent Transcription in Transgenic Zebrafish 57

Chapter Three: A GFP-Dependent Cre Recombinase for Retrofitting Transgenic GFP Lines

Figure 3.1. A GFP-dependent Cre recombinase 78

Figure 3.2. GFP directly controls Cre recombination in a cell-specific manner *in vivo*....80

Figure 3.3. Quantification summary of data from Figure 3.2.....82

Figure 3.4. CRE-DOG appears to act via independently of protein splicing. 85

Figure 3.5. Truncation of N-VMA intein fragment promotes enhanced GFP-specificity and activity of CRE-DOG.....87

Figure 3.6. Characterization of the optimized CRE-DOG.....90

Figure 3.7. Retrofitting transgenic GFP mouse lines for GFP-dependent Cre recombination of specific cell populations. 93

Figure 3.8. Quantification of electroporation experiment in Figure 3.7.....95

Figure 3.9. Delivery of CRE-DOG^{OPT} to the mouse nervous system by rAAV.....99

Chapter Four: Detection and Manipulation of Antigen-Expressing Cells with Destabilized Binding Proteins

Figure 4.1. Concept of protein-inducible protein stabilization 113

Figure 4.2. Isolation of a highly destabilized GBP1 variant that can be stabilized in expression by GFP 115

Figure 4.3. dGBP1 can confer GFP-dependency to multiple fusion protein partners. 118

Figure 4.4. Effect of dGBP1 multimerization on background recombinase activity.....121

Figure 4.5. Mapping of key destabilizing mutations in dGBP1 123

Table 4.1. Semi-quantitative summary of results from dGBP1 destabilizing residue mapping experiments 125

Figure 4.6. GFP can control protein stabilization in a cell type-specific manner *in vivo* 127

Figure 4.7. Structural alignment of three antigen-nanobody complexes. 130

Figure 4.8. Destabilized CTD nanobody generated by grafting the destabilizing mutations from dGBP1. 132

Chapter Five: Concluding Discussions

Figure 5.1. Summary of GFP-dependent systems developed in this thesis 150

Figure 5.2. Co-opting endogenous intracellular products for cell-specific gene manipulation 154

Appendix II: Rejected Cover Art

Figure 1. Rejected Cell Cover 1.....174

Figure 2. Rejected Cell Cover 2.....175

Acknowledgements

This thesis would not have been possible without the unwavering support of my mentor Connie Cepko. Connie is amongst a dying breed of scientists who went after interesting problems without worrying about where the results would be published. She gave me the freedom to tap into my creative side that probably would have been buried by a more "focused" PI. Perhaps the most valuable lessons came from observing how Connie deals with adversity; I am still trying to learn from her ability to skillfully resolve conflicts and crises.

I am also grateful to my DAC members, Kevin Struhl, Michael Springer and Jesse Gray for the advices throughout the process, as well as examiners Martin Chalfie, Gary Yellen and Mark Andermann for participating in my dissertation defense. I would also like to thank past and present Cepko/Tabin/Dymecki lab members for all their feedback and support regarding my projects over the years. Thanks to Kim Burman, Vonda Shannon, Ella Sexton and Terri Broderick for helping out with all sorts of bureaucratic stuff.

Amongst the many people who deserve mention in the Tri-lab, I would like to thank Mark Emerson, Nathan Billings, Kevin Beier, Claudio Punzo, Rahul Kanadia, Jerome Gros, Russell Ray, Akinori Kan, Jessica Lehoczky, Jessica Whited for helping me pick up various techniques and analysis approaches early on in my time here. It is also nice to have stimulating discussions with various members over the years, especially Alan, Mark, Sylvain, Gene, Nathan Mundell and Yash. It is always great to chat with the cheerful Wenjun. I feel like I grew up alongside Sui, Wenjia and Natalia the most as we

all started around the same time in the lab. I will miss all the outings I have with the later graduate students: Brian, Cem, Ariel and Tyler for coffee, beer and food.

Thanks to Friday Harbor Laboratories, Billie Swallie and Claudia Mills for giving me the opportunity to experience what it was like to be Osamu Shimomura hunting for jellyfishes at the docks. Thanks to all those associated with the MBL Neurobiology course at Woods Hole for 2 months of wonderful memories. I feel especially lucky to have met a wonderful classmates in Tan (thanks for all the humor!)

I would like to acknowledge my former mentor Douglas Allan for continuing to support me over the years. I would also like to thank Jacob Hodgson for chatting with me over the years about my various issues. Thanks to Ben Huang and Jonathan Coleman for being amazing friends along the process.

Much love to my parents Joe and Mary for supporting me unconditionally, and to Kailun for joining me for the ups and downs of this roller coaster ride called grad school.

Dedication

To the Canada Geese of Emerald Necklace

for all the free entertainment and lessons on animal lifestyle.

Chapter One

Introduction

A reductionist approach to dissecting complex multicellular systems

The construction and function of complex multicellular organisms involves the action of phenotypically distinguishable "cell types" that specialize in different tasks. Thus, one way to dissect mechanisms underlying specific developmental, physiological or behavioral phenomena is to refine experimental manipulations to individual cell types. This simple reductionist approach has practical challenges. One may appreciate these challenges by examining the mammalian retina, the nerve tissue that lines the back of the eye and is involved in vision. Decades of anatomical studies have revealed the existence of around 50 morphologically distinct retinal cell types, each either known or expected to participate uniquely in the scheme of visual processing (Masland, 2001) (Figure 1.1). Yet, despite this knowledge we are only in the beginning stages of understanding how individual cell types contribute to visual processing via their neural activity and connectivity. This is because existing technologies and resources limits the extent to which we can target and manipulate any desired cell type for experimentation. In this thesis, I introduce concepts and technologies for facilitating the control of specific cell types across nervous systems and other multicellular systems.

Defining cell types

Since the drawings of Ramon y Cajal (Cajal, 1911) established the neuron as the basic unit of the nervous system, it is recognized that an understanding of how the brain works would involve defining the properties and function of individual neuronal cell types. So what is a neuronal cell type? A cell type is a classification used to treat a population of cells as a homogenous unit, based on phenotypically definable traits. A cell type can be

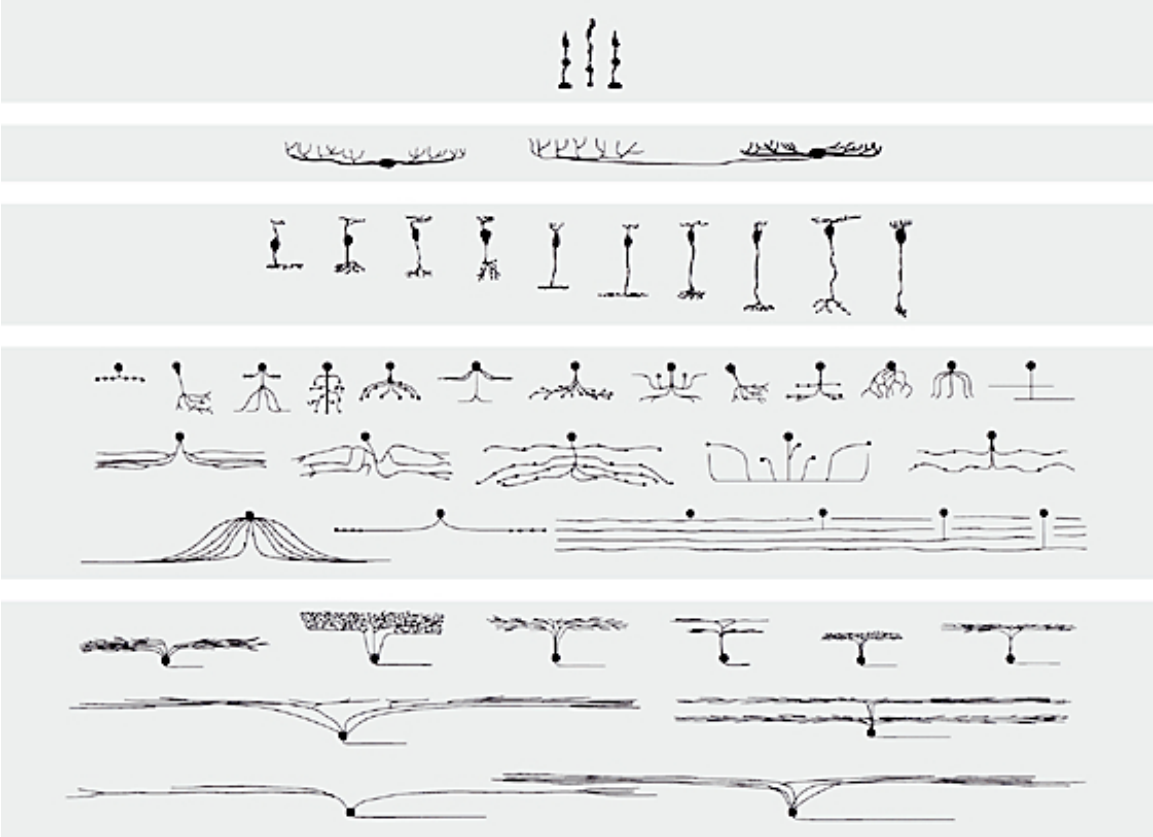


Figure 1.1. Diversity of cell types in the typical mammalian retina

Depicted cell types are based on data from a number of species. The major retinal cell classes are grouped into grey boxes. The cell classes are, from top to bottom: photoreceptors, horizontal cells, bipolar cells, amacrine cells, ganglion cells. Adopted schematic (Masland, 2001)

categorized based on many observable phenotypes, ranging from gene expression, axonal projections, morphologies, cell body location, neuronal activity, etc. Since this classification is arbitrary, specific cell types classified based on one trait may often be further sub-divided into multiple "sub-types" based on additional traits. Of course, increasingly precise distinction between cells may involve combined number of traits. However, the precision to which cell types are classified is currently limited by our ability to analyze the many phenotypic traits that exist amongst cells.

Two cells of the same type may or may not be functionally identical. For example, cells expressing the VGluT1 gene may be referred to as excitatory neurons based on VGluT1's ability to support glutamate release at synaptic terminals and thereby depolarize postsynaptic terminals, but sub-populations of VGluT1+ cells may project to entirely different targets and the excitatory action of these subtypes may exert different effects in the neuronal network, resulting in differences in brain activity and behavior.

How does one target and experiment on a single cell type? By exploiting the very traits that make cell types unique. The focus of this discussion will be on genetic approaches to target genes that are selectively expressed in specific cell types. Currently, the predominant approach in the fields of developmental biology and neuroscience is to co-opt the *cis*-regulatory elements that activate transcription of a cell type-specific gene of interest (Luo et al., 2008) (Figure 1.2). By putting a gene of choice under the transcriptional control of specific cis-regulatory elements, one may be able to express desired genes in the cell type of interest. While this approach is highly successful and

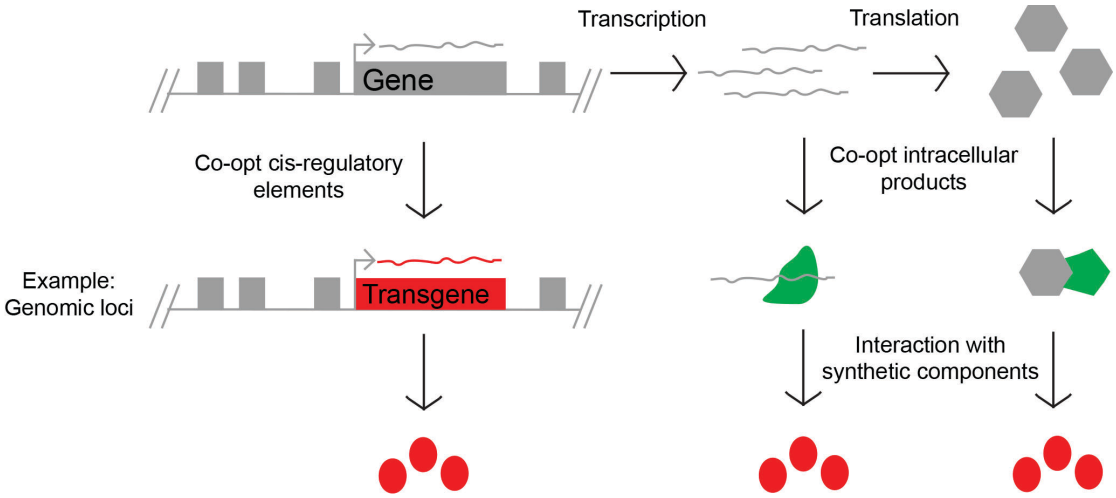


Figure 1.2. Targeting cells by exploiting endogenous processes. To target a cell type based on unique molecular profile, one can co-opt the cis-regulatory elements regulating a cell-specific gene of interest. In the example shown, one might insert a transgene (red) into the endogenous loci of the gene of interest (gray), resulting in expression of the transgene products (red ellipses). Alternatively, one can target the transcribed and translated products from the endogenous loci. Interactions between genomic products with synthetic components (green shapes) leads to the expression of the desired transgene product (red ellipses).

simple to implement, a problem with this approach is that one needs to know the cis-regulatory sequences to be used for targeting a specific cell type. The other problem is that when these cis-regulatory sequences are placed in random locations within a genome, position effects exerted by neighboring elements may alter the expression pattern of a cis-regulatory element of interest, resulting in undesired expression of the transgene (Feng et al., 2000).

Alternatively, one may target cell types by co-opting the molecules that are uniquely expressed or present in cells of interest. One way this is accomplished is by the delivery of cargo carrying vectors, such as cell-penetrating peptides, liposomes, viruses, DNA origami, etc, that target cells based on their expression of specific cell surface molecules (Douglas et al., 2012; Federspiel et al., 1994; Mallick and Choi, 2014; Paszko and Senge, 2012; Svensen et al., 2012). In one example, viruses with specific tropism for a cell type-specific surface molecule would only be taken up by these cells of interest, resulting in expression of the transgene selectively amongst other cell types (Federspiel et al., 1994). Another way one might envision targeting specific cell types is to exploit the unique repertoire of intracellular products expressed in a cell (Figure 1.2). This approach has only recently become possible, owing to the development of novel reagents and approaches for binding and co-opting intracellular products. The ability to co-opt endogenous protein as gene-inducing signals was demonstrated with an RNA controller whereby an aptamer that binds to the endogenous protein is inserted into an engineered intron placed upstream of a target gene of interest. Binding of the endogenous protein to the aptamer results in changes to mRNA splicing outcome, resulting in exclusion of a

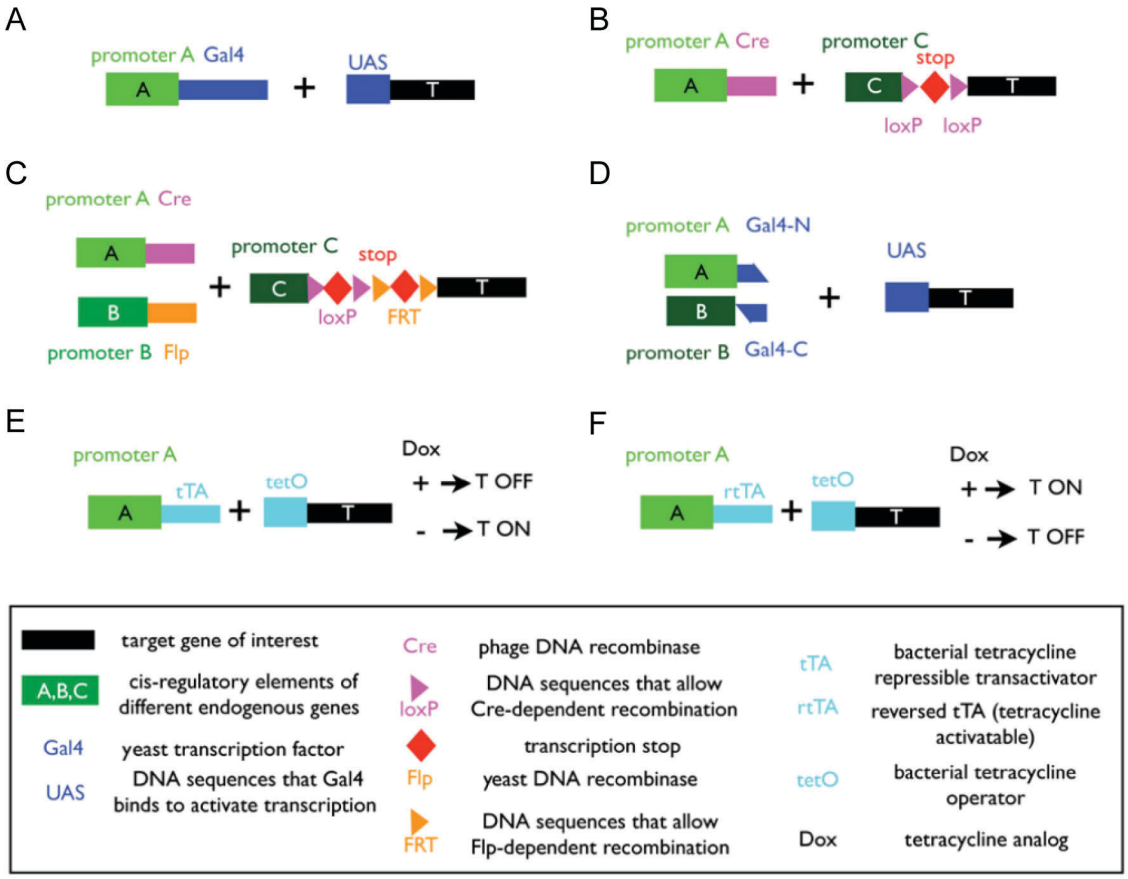
disruptive exon and in the increased expression of a target gene of interest (Culler et al., 2010). This promising strategy awaits application in animals (Chang et al., 2012). In this thesis, I will discuss the use of another type of reagent, the nanobody, for co-opting intracellular proteins for cell-specific gene manipulation in animals.

Genetic tools facilitating the study of cell types

With genetic access to a cell type of interest, it becomes possible to control the properties of the cell by expressing genes of choice. One may accomplish this by directly fusing coding sequences of interest to the cell type-specific promoters of interest. This becomes an increasingly complicated approach since it is necessary to fuse the many different gene products to many different cell type-specific promoters. Further, the activity of gene products may be strongly influenced by the expression of the promoter to which it is fused to, making it difficult to get a gene to high enough level for desired effects. One solution to these problems is to make the cell-type specificity a modular component that can be combined with the gene output component as a separate module (Figure 1.3). For example, one may express an exogenous transcription factor under the control of a cell-type specific promoter, upon which the transcription factor can then bind to the upstream activating sequences (UAS) of a gene of interest, resulting in activation of the desired gene products. This is exemplified by the GAL4/UAS system most successfully applied in Drosophila (Brand and Perrimon, 1993) (Figure 1.3A). GAL4 encodes a yeast transcription factor Gal4, which can bind to the UAS elements, resulting in transcription of the gene placed downstream of the UAS-promoter. Related to

Figure 1.3. Binary systems and strategies commonly used to study genetically tractable model organisms. (A-B) Combining Gal4 (A) or Cre (B) expression with a UAS- (A) or loxP-based (B) responder cassette leads to expression of target gene (T) in cells that are actively transcribing from promoter A. In (B), the target gene is inactive until Cre removes the transcriptional stop cassette that is flanked by loxP sequences. Promoter C is usually broadly active across cells, but can be cell-specific as well. (C) Variation of the binary strategy allows for increasingly precise cell-specific manipulations. In this case cell-specific expression of Cre under promoter A is intersected with cell-specific expression of Flp under promoter B, leading to expression of target gene only in cells active in A and B. (D) An alternative intersectional strategy. Here, Gal4 has been split into two pieces, and only when the two pieces are co-expressed in a cell type can the target gene be expressed. (E-F) Temporal control of binary systems. Here, the Tetracycline transactivator (tTA) or reverse tTA (rtTA) is dependent on the presence of drug Doxycycline (Dox). Dox inhibits DNA binding of tTA (E) and promotes DNA binding of rtTA (F). DNA binding at the tetO responder element leads to transcription of target gene. Symbols explained in the boxed legend. Adopted schematic (Luo et al., 2008).

Figure 1.3. (Continued)



GAL4/UAS system are other transcription factor/promoter systems such as LexA/lexAop (Butala et al., 2009) and TetON/OFF systems (Figure 1.3E-F) (Schonig et al., 2010). In another example, a site-specific DNA recombinase such as Cre, from bacteriophage p1, and Flp, from yeast, are expressed off cell-specific promoters and used to induce DNA recombination events via recognition of their respective cognate binding sequences loxP and FRT (Dymecki, 1996; Orban et al., 1992) (Figure 1.3B). The advantage of such "binary" systems is that one can easily mix and match cell-type specific "drivers" with "responder" cassettes. For example, if there were 10 cell-type specific promoters and 10 genes to be expressed per promoter, the number of direct promoter-gene fusion constructs would be 10x10=100 constructs, whereas with a binary system the number of constructs to be made would be 10+10=20. This becomes significant when the constructs are made into transgenic animals.

Many definable cell types cannot be labeled by a single gene marker (Dymecki et al., 2010). To facilitate the study of cell types at increased cellular resolution, one can arrange for the expression of a target gene to be dependent on the activity of multiple cell-specific promoters. The intersection of these cell-specific signals could be mediated by a combination of driver molecules converging on a single responder cassette (Figure 1.3C), or by splitting a single driver molecule into multiple components, and driving the separated parts from different cell-specific promoters (Figure 1.3D).

Green Fluorescent Protein

Around two decades ago, the demonstration that the Green Fluorescent Protein (GFP) (Shimomura et al., 1962), from the jellyfish *Aequorea victoria*, could be used as a marker of gene expression triggered a revolution in biology (Chalfie et al., 1994) (Figure 1.4). GFP is a beta barrel structure with a chromophore in the middle, formed by a spontaneous maturation process resulting in cyclization and oxidation of a tri-peptide Ser-Tyr-Gly (Figure 1.4C) (Tsien, 1998). Because GFP fluoresces without the need for cofactors, and could be conveniently visualized under fluorescence microscopes, it has become a pervasive tool for visualizing the many cell types of the nervous system and other tissues (Figure 1.4D). Indeed, GFP has since been shown to be a useful gene expression marker across model organisms, from bacteria to worms to flies to fish to mice and even primates and plants (Figure 1.4D) (Chalfie, 2009). Thousands of transgenic animals have now been generated that use GFP as a reporter of cell-specific expression of genomic fragments, endogenous loci, or defined cell-specific promoter elements. Notably, in the mouse, over 1500 transgenic GFP reporter lines have been established, with a predominant approach employing ~200kb genomic fragments encoded in Bacterial Artificial Chromosomes (BAC) as constructs for transgenic lines (Figure 1.5A) (Gong et al., 2003; Heintz, 2004). Many of the GFP expression patterns have been described and listed in the public domain (Figure 1.5B-C) (gensat.org) (Siegert et al., 2009).

Figure 1.4. The development of Green Fluorescent Protein as a tool. (A) The Green Fluorescent protein (GFP) was originally isolated from *Aequorea* species of jellyfish at the docks of Friday Harbor Laboratories in Friday Harbor, Washington. (B) GFP is expressed in a subset of cells positioned along the rim of the umbrella structure, and releases green fluorescence when disrupted mechanically (right panel). (C) The crystal structure of GFP showing a beta-barrel structure, with the tri-peptide chromophore shown inside the barrel. (D) GFP is now commonly used as a marker of gene expression across model systems in biology (E) Protein engineering resulted in the a palette of fluorescent proteins emitting at different wavelengths along the visible spectrum. Adopted images from nobelprize.org.

Figure 1.4. (Continued)

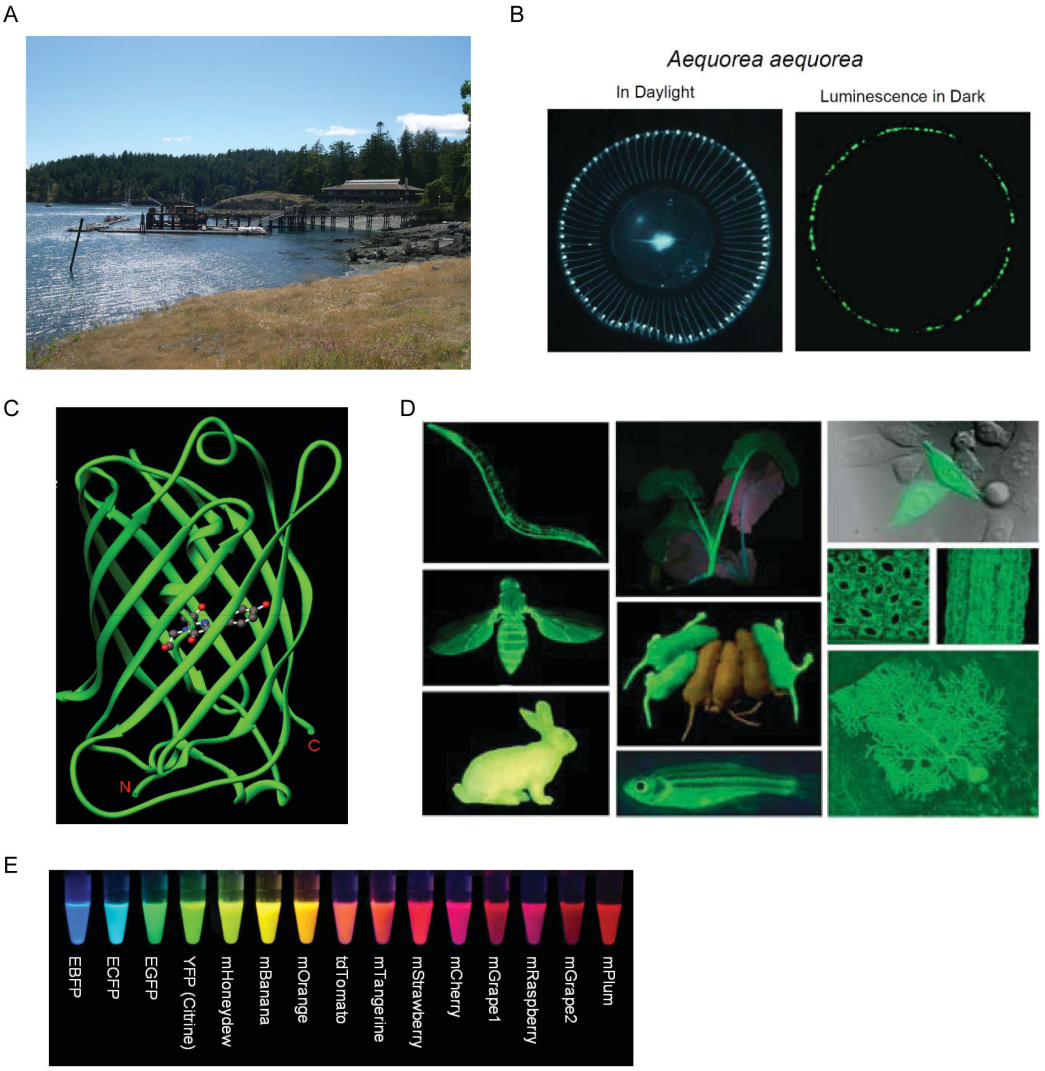
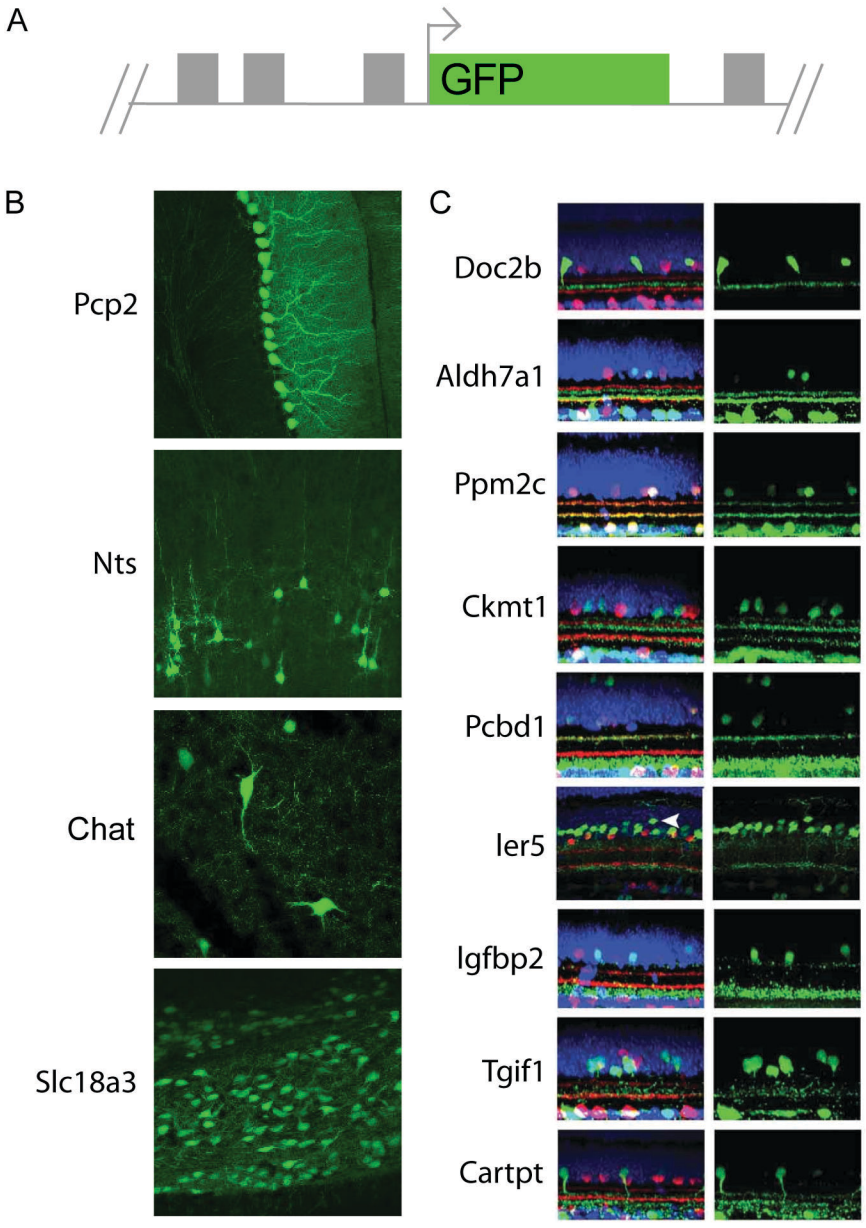


Figure 1.5. Diverse transgenic GFP reporter lines have been generated and characterized for the mouse community. (A) A systematic effort to generate transgenic GFP reporter lines has been taken in the mouse community. This focuses on utilizing Bacterial Artificial Chromosome (BAC) to construct GFP reporter constructs under the control of large ~250 kbp genomic fragments. (B-C) A sample of the many neuronal cell types labeled by different GFP lines in the brain (B) and retina (C). The red marker in (C) is a Anti-Chat immunostain that marks specific layers of the inner plexiform layer of the retina, used to align the axonal processes of GFP cell types labeled in different lines. Adopted images in (B) from gensat.org. Adopted figures in (C) (Siegert et al., 2009)

Figure 1.5. (Continued)



Limitation of existing GFP tools, and approach to cell-specific gene manipulation

Despite the widespread application of GFP, its potential uses remain underexplored. One limitation of GFP is that it currently has no known regulatory abilities; GFP can be used as a fluorescent marker in cells, but it cannot be used to control biological activities. The ability to use GFP as a regulatory molecule inside cells would be significant, because one could then make use of the thousands of transgenic GFP reporter lines as tools for cell-specific gene manipulation studies, without having to create new driver lines expressing other driver molecules such as Cre or GAL4. In the mouse community this would be especially useful, because hundreds to thousands of transgenic GFP expression patterns have been characterized in the nervous system, and transgenic lines can be obtained from a public depository. Further, given the long generation time and cost associated with generating transgenic mouse lines, the ability to immediately make use of existing GFP reporter lines for experimental manipulations of desired cell populations would speed up the rate of discovery. In fact, at the beginning of my Ph.D., I realized that the mouse community had a collection of transgenic GFP reporter lines that labeled the class of retinal cell types, the bipolar cells, of my interest, but we were lacking the transgenic driver Cre lines needed at the time to control gene expression selectively in the GFP-labeled cell types (Figure 1.10B). One solution would be to simply take the same genomic fragment that drove GFP expression in labeled cell types and create a new transgenic mouse expressing Cre. However, this approach is time-consuming and costly, and could potentially suffer from changes to the expression pattern of Cre due to position effect exerted by surrounding loci at the random integration site. To bypass this issue, I proposed to directly make use of transgenic GFP reporter lines as tools for cell-specific

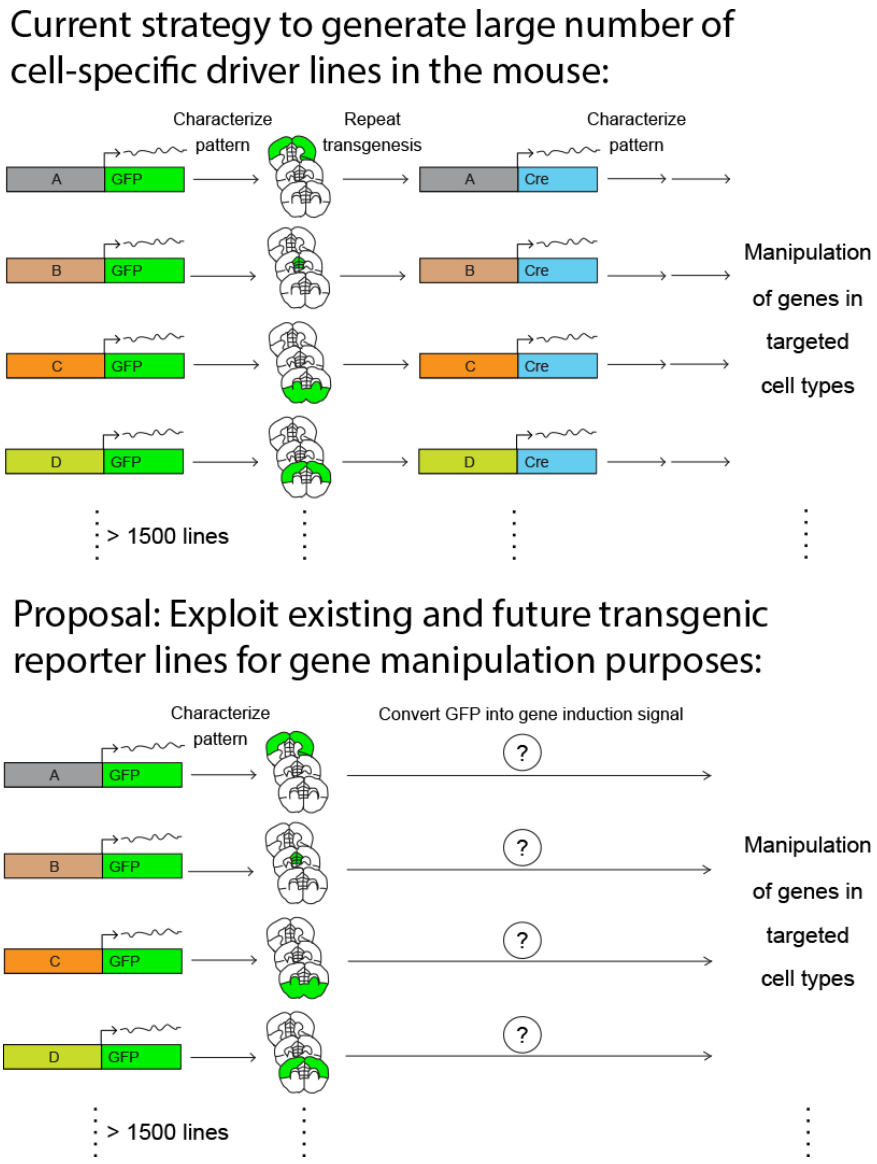
gene manipulation, by creating synthetic systems that can interface with GFP expression and convert that interaction into desired molecular outputs. This was the aim of my thesis (Figure 1.6), which further led to new perspectives on cell-specific manipulations.

Nanobody reagents for use with Green Fluorescent Protein

To create synthetic systems controllable by GFP, one needs molecular parts that can engage with GFP. GFP is a relatively innocuous molecule that can be easily applied across model organisms. Part of the reason may be due to its lack of interactions with host proteins across diverse cell types and across species. When the GFP sequence is expressed without any tags, it is cytoplasmic and can enter the nucleus, probably because its size (~25kDa) permits passive diffusion through the nuclear pore complex. The lack of GFP interaction with host cell proteins is confirmed in at least one mass spectrometry experiment in mammalian cell culture (Trinkle-Mulcahy et al., 2008). The lack of connectivity to host interactome makes GFP highly suitable for building synthetic circuits around, by way of creating protein-protein interactions that now integrate the molecule into the interactome. Reagents targeting the GFP molecule have recently been developed. One class of this is GFP binding proteins (GBP) (Figure 1.7) (Kirchhofer et al., 2010; Rothbauer et al., 2008; Rothbauer et al., 2006). These are antigen-recognition domains, or nanobodies, taken from the antibodies of Camelid species (Muyldermans, 2013). The unique feature of Camelid antibodies is that the antigen-recognition domain is encoded in a single polypeptide chain as opposed to separated polypeptides seen in conventional antibodies (Figure 1.8A). Camelid antibodies hold the advantage that the nanobodies can be easily expressed as a monomer in living cells, in a manner that allows

Figure 1.6. Exploiting transgenic reporter lines as tools for cell-specific gene manipulation. Top, conventional strategy for generating large number of cell-specific Cre driver mouse lines. Large genomic fragments (colored and labeled A-D) are used to drive GFP expression in various cell populations across transgenic lines. Transgenic GFP expression patterns inform the selection of specific genomic fragments for creating additional transgenic lines expressing Cre recombinase, which is used for gene manipulation. Bottom, proposal to directly make use of transgenic GFP reporter lines for cell-specific gene manipulation. If it is possible to use GFP to control gene expression, then one can bypass the need to generate redundant Cre-expressing lines using the same genomic fragments.

Figure 1.6. (Continued)



for solubility of the protein as well as retain affinity for the antigen. A number of GBP has been isolated (Kirchhofer et al., 2010). Notably, GBP1 and GBP4 bind GFP in such a way as to either enhance or minimize GFP fluorescence, respectively (Figure 1.7A)(Kirchhofer et al., 2010).

Notable features of nanobodies

Nanobody primary sequence can be classified into 3 hypervariable regions (HV) and into 4 conserved framework regions (FR) (Figure 1.8B). The polypeptide forms 9 beta strands and when folded into a tertiary structure each of the 3 HV regions form loops between beta strands to form a surface along one end of the structure to facilitate epitope recognition. These regions are called the complementarity-determining regions (CDR) (Figure 1.8B). Several types of antigen recognition surface, or paratope, have been described for nanobodies (Muyldermans, 2013). Much of the difference involves changes to the loop structure corresponding to CDR3. Most paratope surfaces involve folding of the CDR3 loop into a flat surface over FR2. However, in the case of the GFP binding protein 1 (GBP1) nanobody, which will be the focus of Chapter 4, the CDR3 region is relatively shorter than most of their counterparts. These differences become relevant for the work in Chapter 4.

Strategies for controlling biological activities with proteins – concepts related to small molecule regulated systems

How would one make use of an intracellular product for cell-specific control? The key is to design synthetic systems that become active upon interaction with the product. The

Figure 1.7. GFP-binding nanobodies. (A) A panel of GFP binding nanobodies tested for ability to affect GFP fluorescence *in vitro*. GBP1 has fluorescence enhancing effects while GBP4 has fluorescence minimizing effects. (B-C) Crystal structures showing GBP1 (purple structure in B) and GBP4 (orange structure in C) bound to GFP (green). Adopted figures and images (Kirchhofer et al., 2010).

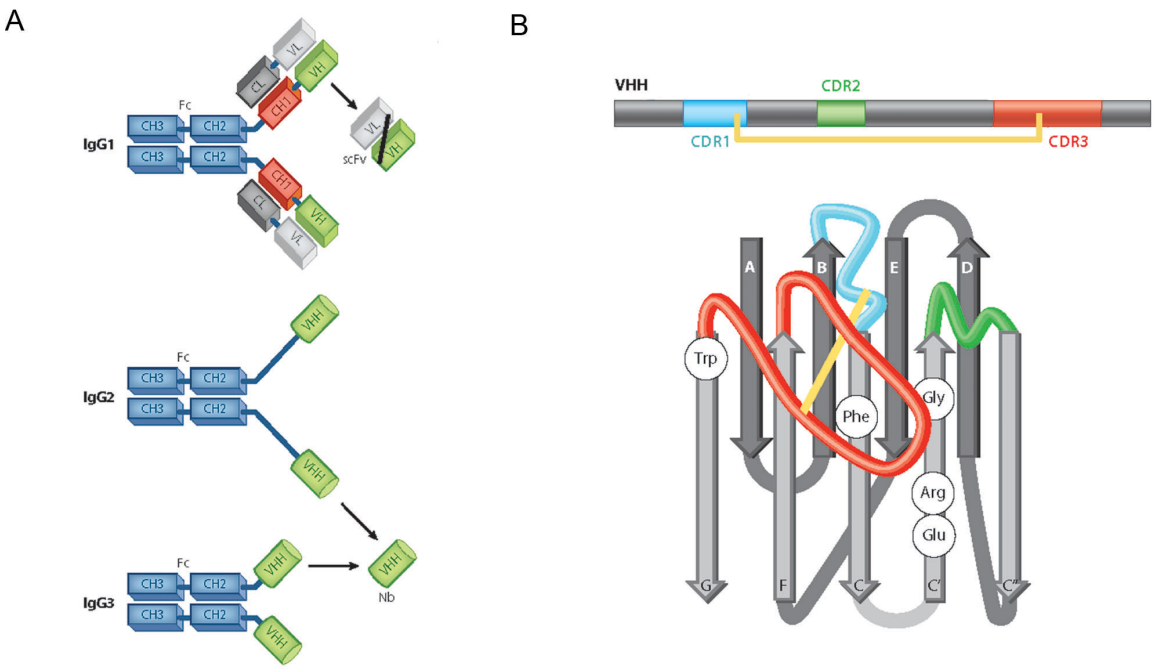


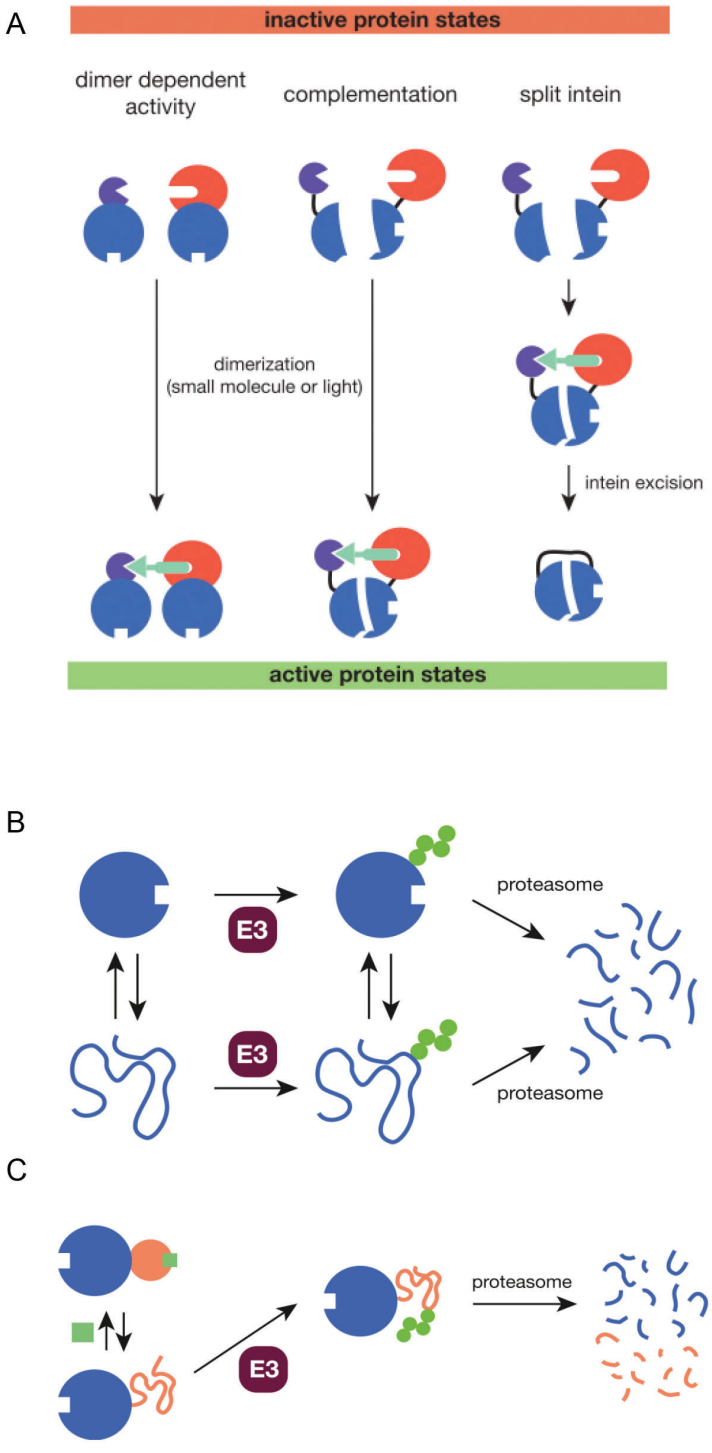
Figure 1.8. Camelid antibodies and nanobodies. (A) Camelid species have two types of immunoglobulins (IgG). IgG1 is the conventional type, with heavy and light chain joined together by disulfide bonds to form the antigen recognition pocket. IgG2 and IgG3 are unique from most other species in that they contain only the heavy chain, and the antigen recognition domain (VHH) is encoded in a single polypeptide. When isolated, they are called nanobodies (Nb). (B) Primary structure of VHH coding sequence (top). FR, framework. CDR, complementarity determination region. Tertiary structure of VHH (bottom). Displayed residues are sites that would have interacted with the light chain in conventional antibodies, but are now mutated to better accommodate the new environment. Adopted schematic (Muyldermans, 2013).

work in this thesis borrows from concepts that have been primarily used to design small molecule regulated systems, intended for temporal control. In one scheme, a small molecule is used as a "dimerizer" to bring together two complementary protein domains, resulting in reconstitution of desired biological activity (Figure 1.9A). One example involves hybrid transcription factors, whereby the DNA binding domain and the transactivation domains are assembled onto a promoter of interest as targeted by the DNA binding specificity (Pollock and Clackson, 2002). Another example concerns split protein enzymes such as site-specific DNA recombinase Cre. (Jullien et al., 2003). One further process that dimerizers can regulate is protein splicing. Protein splicing is a process whereby two protein regions, or exteins, are covalently joined together by a catalysis mediated by an intervening protein domain called inteins. In a dimerizer-regulated scenario, the two exteins are expressed as separate fragments fused to split portions of an intein. Dimerization induces association and activation of the split inteins, resulting in protein splicing and formation of a single polypeptide chain joining the two exteins (Figure 1.9A) (Mootz and Muir, 2002; Tyszkiewicz and Muir, 2008).

Small molecules have also been used to control protein stability. In such a system, a mutated small molecule-binding domain is rendered inactive or highly degradable, resulting in absence or low levels of the domain. Exposure to the small molecule binder stabilizes the domain, leading to accumulation of whatever protein of interest is fused to the destabilized domain. (Banaszynski et al., 2006; Bonger et al., 2014). In this thesis, I will explore using both the dimerizer (Chapter 2,3) and destabilizing domain (Chapter 4) strategies to confer intracellular proteins with synthetic regulatory abilities.

Figure 1.9. Two possible strategies exploiting intracellular proteins as signals to control gene expression. (A) Dimerizer-regulated strategy. A dimerizer brings together proteins fragments or modular domains that are inactive for the desired biological activity, but become biologically active upon formation of a dimerizer-induced complex. In the case of split intein, a functional protein is generated by covalently joining the two fragments together into one piece, liberating them from dimerizer control. (B-C) Controlling protein activity via ubiquitylation. (B) Damaged or misfolded proteins are targeted for degradation by the proteasome via attachment of four ubiquitin moieties. The ubitquitin E3 ligases handle ubiquitin attachment. (C) Ligand-inducible degradation. A destabilizing domain fused to a protein with desired activity may targets the entire fusion protein for degradation. Upon binding to the ligand, the destabilizing domain is stabilized, leading to accumulation of the fusion protein. Adopted schematics (Rakhit et al., 2014)

Figure 1.9. (Continued)



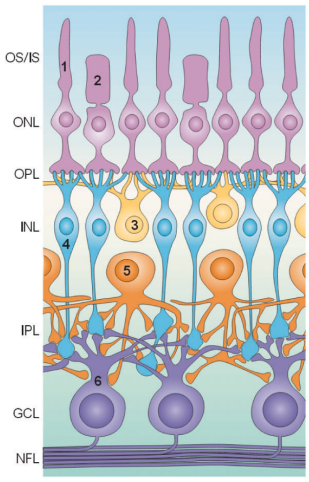
Retina as a model system for testing engineered protein-responsive constructs

As the ultimate aim of synthetic constructs is to facilitate studies of animals, especially the mouse, it is desirable to test the activity of these constructs in a rapid and convenient way *in vivo*. The retina serves as an excellent platform for this purpose, as explained below.

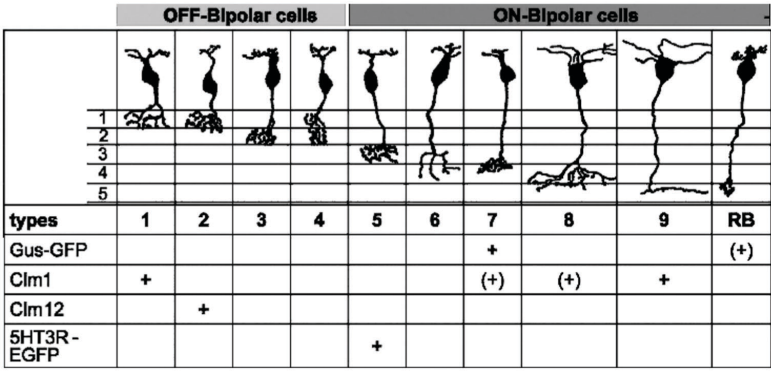
Anatomically, the mouse retina can be divided into three cell layers, the outer plexiform layer (ONL), the inner plexiform layer (INL) and the ganglion cell layer (GCL) (Figure 1.10A). Rod and cone photoreceptors inhabit the ONL, whereas mixtures of horizontal cells, bipolar cells, Muller glial cells as well as amacrine cells inhabit the INL. Ganglion cells and displaced amacrine cells populate the GCL.

Five major cell classes inhabit the retina. In the ONL, rod and cone photoreceptors sense dim light and color light, respectively, and transmit these to the INL, where bipolar cells conveys the signal. Horizontal cells are also connected to photoreceptors. Bipolar mediated signals are then filtered and processed by amacrine cells and ultimately conveyed to ganglion cells, and onto the brain. Within these cell classes lies an additional layer of complexity. In particular, many subtypes of bipolar, amacrines and ganglion cells exist, but the function of each subtype in the scheme of visual processing is still poorly understood (Masland, 2012). A key to overcoming this lack of knowledge is development of novel genetic strategies to turn on functional tools for gene and neural circuit manipulations in specific cell types.

A



B



types	1	2	3	4	5	6	7	8	9	RB
Gus-GFP							+			(+)
Clm1	+						(+)	(+)	+	
Clm12		+								
5HT3R-EGFP					+					

Figure 1.10. Organization of the retina and diversity of GFP lines labeling specific cell types. (A) Cross-section of the retina showing the laminated organization of the diverse neuronal cell classes, including rod photoreceptors (1), cone photoreceptors (2), horizontal cells (3), bipolar cells (4), amacrine cells (5), ganglion cells (6). Muller cells, a type of glial cell in the retina, is not depicted. (B) Within the bipolar cell class a diversity of morphologically distinct subtypes exist. The rationale for this thesis came from the realization that multiple transgenic GFP lines were already found to label specific bipolar subtypes. Many more GFP lines in the GENSAT collection and else where (not shown here) also label a wide variety of retinal subtypes. Adopted schematic (Wassle, 2004; Wassle et al., 2009).

A number of reasons make the retina suitable as a testbed for synthetic constructs *in vivo*. First, it is part of the central nervous system, and thus can be a proxy for testing whether constructs will work in the brain. Second, electroporation methods have been established for delivery of multiple DNA plasmids into cells during development. This is a significant advantage because it is fairly non-invasive to inject DNA into the postnatal pup, compared to the surgical procedures used for in utero electroporation into the developing embryonic brain. Third, the retina is a well-laminated structure and broad cell classes and even specific subtypes can be identified by their position in the laminated structure as well as by morphology. This ensures an easy way to analyze the distribution of cell composition labeled by electroporation. (Matsuda and Cepko, 2004, 2007).

Summary of thesis

In this thesis, I invented novel synthetic systems for repurposing GFP as a functional tool for gene manipulations. In Chapter 2, I present my first attempt. I took advantage of GFP binding proteins derived from Camelid antibodies to build a GFP-dependent transcription system, which utilizes GFP to bring modular transcription domains together into an active transcription factor. I demonstrated the utility of GFP as a useful tool for manipulating specific cell populations in the mouse nervous system using this method.

To overcome caveats associated with the GFP-dependent transcription system, I further explore the use of GFP to directly control site-specific DNA recombination. In Chapter 3, I developed a system whereby GFP brings together split Cre fragments for DNA recombination. I show the utility of this system in the mouse. In Chapter 4, I simplified

the use of GFP by using GFP to stabilize the expression of a protein of interest. The strategies described in this thesis have implications beyond GFP, as it sets the stage for generation of protein-responsive sensors and effectors for cell-specific interrogations and manipulations in wildtype animals.

Reference

Banaszynski, L.A., Chen, L.C., Maynard-Smith, L.A., Ooi, A.G., and Wandless, T.J. (2006). A rapid, reversible, and tunable method to regulate protein function in living cells using synthetic small molecules. Cell *126*, 995-1004.

Bonger, K.M., Rakhit, R., Payumo, A.Y., Chen, J.K., and Wandless, T.J. (2014). General method for regulating protein stability with light. ACS Chem Biol *9*, 111-115.

Brand, A.H., and Perrimon, N. (1993). Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. Development *118*, 401-415.

Butala, M., Zgur-Bertok, D., and Busby, S.J. (2009). The bacterial LexA transcriptional repressor. Cell Mol Life Sci *66*, 82-93.

Cajal, S.R. (1911). Histology of the nervous system of man and vertebrates, 1995 translation edn (Oxford, Oxford University Press, Inc).

Chalfie, M. (2009). GFP: Lighting up life. Proc Natl Acad Sci U S A *106*, 10073-10080.

Chalfie, M., Tu, Y., Euskirchen, G., Ward, W.W., and Prasher, D.C. (1994). Green fluorescent protein as a marker for gene expression. Science *263*, 802-805.

Chang, A.L., Wolf, J.J., and Smolke, C.D. (2012). Synthetic RNA switches as a tool for temporal and spatial control over gene expression. Curr Opin Biotechnol.

Culler, S.J., Hoff, K.G., and Smolke, C.D. (2010). Reprogramming cellular behavior with RNA controllers responsive to endogenous proteins. Science *330*, 1251-1255.

Douglas, S.M., Bachelet, I., and Church, G.M. (2012). A logic-gated nanorobot for targeted transport of molecular payloads. Science *335*, 831-834.

Dymecki, S.M. (1996). Flp recombinase promotes site-specific DNA recombination in embryonic stem cells and transgenic mice. Proc Natl Acad Sci U S A *93*, 6191-6196.

Dymecki, S.M., Ray, R.S., and Kim, J.C. (2010). Mapping cell fate and function using recombinase-based intersectional strategies. Methods Enzymol *477*, 183-213.

Federspiel, M.J., Bates, P., Young, J.A., Varmus, H.E., and Hughes, S.H. (1994). A system for tissue-specific gene targeting: transgenic mice susceptible to subgroup A avian leukosis virus-based retroviral vectors. Proc Natl Acad Sci U S A *91*, 11241-11245.

Feng, G., Mellor, R.H., Bernstein, M., Keller-Peck, C., Nguyen, Q.T., Wallace, M., Nerbonne, J.M., Lichtman, J.W., and Sanes, J.R. (2000). Imaging neuronal subsets in transgenic mice expressing multiple spectral variants of GFP. Neuron *28*, 41-51.

Gong, S., Zheng, C., Doughty, M.L., Losos, K., Didkovsky, N., Schambra, U.B., Nowak, N.J., Joyner, A., Leblanc, G., Hatten, M.E*., et al.* (2003). A gene expression atlas of the central nervous system based on bacterial artificial chromosomes. Nature *425*, 917-925.

Heintz, N. (2004). Gene expression nervous system atlas (GENSAT). Nat Neurosci *7*, 483.

Jullien, N., Sampieri, F., Enjalbert, A., and Herman, J.P. (2003). Regulation of Cre recombinase by ligand-induced complementation of inactive fragments. Nucleic Acids Res *31*, e131.

Kirchhofer, A., Helma, J., Schmidthals, K., Frauer, C., Cui, S., Karcher, A., Pellis, M., Muyldermans, S., Casas-Delucchi, C.S., Cardoso, M.C*., et al.* (2010). Modulation of protein properties in living cells using nanobodies. Nat Struct Mol Biol *17*, 133-138.

Luo, L., Callaway, E.M., and Svoboda, K. (2008). Genetic dissection of neural circuits. Neuron *57*, 634-660.

Mallick, S., and Choi, J.S. (2014). Liposomes: versatile and biocompatible nanovesicles for efficient biomolecules delivery. J Nanosci Nanotechnol *14*, 755-765.

Masland, R.H. (2001). The fundamental plan of the retina. Nat Neurosci *4*, 877-886.

Masland, R.H. (2012). The neuronal organization of the retina. Neuron *76*, 266-280.

Matsuda, T., and Cepko, C.L. (2004). Electroporation and RNA interference in the rodent retina in vivo and in vitro. Proc Natl Acad Sci U S A *101*, 16-22.

Matsuda, T., and Cepko, C.L. (2007). Controlled expression of transgenes introduced by in vivo electroporation. Proc Natl Acad Sci U S A *104*, 1027-1032.

Mootz, H.D., and Muir, T.W. (2002). Protein splicing triggered by a small molecule. J Am Chem Soc *124*, 9044-9045.

Muyldermans, S. (2013). Nanobodies: natural single-domain antibodies. Annu Rev Biochem *82*, 775-797.

Orban, P.C., Chui, D., and Marth, J.D. (1992). Tissue- and site-specific DNA recombination in transgenic mice. Proc Natl Acad Sci U S A *89*, 6861-6865.

Paszko, E., and Senge, M.O. (2012). Immunoliposomes. Curr Med Chem *19*, 5239-5277.

Rakhit, R., Navarro, R., and Wandless, T.J. (2014). Chemical Biology Strategies for Posttranslational Control of Protein Function. Chem Biol *21*, 1238-1252.

Rothbauer, U., Zolghadr, K., Muyldermans, S., Schepers, A., Cardoso, M.C., and Leonhardt, H. (2008). A versatile nanotrap for biochemical and functional studies with fluorescent fusion proteins. Mol Cell Proteomics *7*, 282-289.

Rothbauer, U., Zolghadr, K., Tillib, S., Nowak, D., Schermelleh, L., Gahl, A., Backmann, N., Conrath, K., Muyldermans, S., Cardoso, M.C.*, et al.* (2006). Targeting and tracing antigens in live cells with fluorescent nanobodies. Nat Methods *3*, 887-889.

Schonig, K., Bujard, H., and Gossen, M. (2010). The power of reversibility regulating gene activities via tetracycline-controlled transcription. Methods Enzymol *477*, 429-453.

Shimomura, O., Johnson, F.H., and Saiga, Y. (1962). Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, Aequorea. J Cell Comp Physiol *59*, 223-239.

Siegert, S., Scherf, B.G., Del Punta, K., Didkovsky, N., Heintz, N., and Roska, B. (2009). Genetic address book for retinal cell types. Nat Neurosci *12*, 1197-1204.

Svensen, N., Walton, J.G., and Bradley, M. (2012). Peptides for cell-selective drug delivery. Trends Pharmacol Sci *33*, 186-192.

Trinkle-Mulcahy, L., Boulon, S., Lam, Y.W., Urcia, R., Boisvert, F.M., Vandermoere, F., Morrice, N.A., Swift, S., Rothbauer, U., Leonhardt, H.*, et al.* (2008). Identifying specific protein interaction partners using quantitative mass spectrometry and bead proteomes. J Cell Biol *183*, 223-239.

Tsien, R.Y. (1998). The green fluorescent protein. Annu Rev Biochem *67*, 509-544.

Tyszkiewicz, A.B., and Muir, T.W. (2008). Activation of protein splicing with light in yeast. Nat Methods *5*, 303-305.

Wassle, H. (2004). Parallel processing in the mammalian retina. Nat Rev Neurosci *5*, 747-757.

Wassle, H., Puller, C., Muller, F., and Haverkamp, S. (2009). Cone contacts, mosaics, and territories of bipolar cells in the mouse retina. J Neurosci *29*, 106-117.

Chapter Two

A Nanobody-Based System Using Fluorescent Proteins as Scaffolds for Cell-Specific Gene Manipulation

A Nanobody-Based System Using Fluorescent Proteins as Scaffolds for Cell-Specific Gene Manipulation

Authors: Chung Yiu Jonathan Tang[1,2], Tamas Szikra[4], Yevgenia Kozorovitskiy[1,3], Miguel Teixeira[4,5], Bernardo L. Sabatini[1,3], Botond Roska[4], Constance L. Cepko[1,2]

[1]Howard Hughes Medical Institute

[2]Departments of Genetics and Ophthalmology

[3]Department of Neurobiology

Harvard Medical School, Boston, MA 02115, USA

[4]Neural Circuit Laboratories, Friedrich Miescher Institute for Biomedical Research, 4058 Basel, Switzerland

[5]University of Basel, 4003 Basel, Switzerland

This chapter contains the manuscript titled "A Nanobody-Based System Using Fluorescent Proteins as Scaffolds for Cell-Specific Gene Manipulation", published in Cell. It is modified to better fit the style of this dissertation. Jonathan C.Y. Tang is first author on this manuscript, and conceived of GFP-dependent systems. J.C.Y.T, T.S., Y.K., B.L.S., B.R., and C.L.C. designed research. J.C.Y.T. conducted all experiments and analysis except those involving brain electroporation and electrophysiology, T.S., Y.K., and M.T. conducted electrophysiology-related experiments and analysis. B.L.S, B.R. and C.L.C. supervised the project. J.C.Y.T and C.L.C wrote the paper, with contributions from T.S., Y.K., M.T., B.LS. and B.R.

Summary:

Fluorescent proteins are commonly used to label cells across organisms, but the unmodified forms cannot control biological activities. Using GFP-binding proteins derived from Camelid antibodies, we co-opted GFP as a scaffold for inducing formation of biologically active complexes, developing a library of hybrid transcription factors that control gene expression only in the presence of GFP or its derivatives. The modular design allows for variation in key properties such as DNA specificity, transcriptional potency, and drug dependency. Production of GFP controlled cell-specific gene expression and facilitated functional perturbations in the mouse retina and brain. Further, retrofitting existing transgenic GFP mouse and zebrafish lines for GFP-dependent transcription enabled applications such as optogenetic probing of neural circuits. This work establishes GFP as a multifunctional scaffold and opens the door to selective manipulation of diverse GFP-labeled cells across transgenic lines. This approach may also be extended to exploit other intracellular products as cell-specific scaffolds in multicellular organisms.

Introduction:

Studies of multicellular organisms would be greatly facilitated by the ability to manipulate the activities of any genes within specific tissues or cell types. This is challenging to achieve in tissues with diverse cell types, such as the nervous system. (Masland, 2004). To label and provide genetic access to diverse cell types, much effort has been devoted to generating transgenic organisms in which transgenes are placed

under the control of large genomic fragments or endogenous gene loci. Transgenic lines expressing driver genes such as transcription factors or site-specific recombinases can then be used to control the expression of genes in responder cassettes. However, the utility of individual lines is limited by a transgene's functional abilities; reporter lines expressing fluorescent proteins and histochemical enzymes are useful for labeling cells, but cannot currently be used to control biological activities. To replace transgenes driven by the same *cis*-regulatory elements requires generation of additional transgenic lines. Such a procedure can be costly and lengthy for organisms such as the mouse. Thus, a key to conducting efficient and wide-ranging studies on existing and future model organisms is to increase the versatility of transgenic resources.

Owing to their ease of detection, green fluorescent protein (GFP) and its derivatives (Tsien, 1998) have become common markers of gene expression (Chalfie et al., 1994) across model organisms. Notably, thousands of transgenic GFP lines have been generated for the mouse (Gong et al., 2003). This growing and important resource reveals the expression pattern of many genes and provides strains in which GFP selectively labels many cell types of interest (*www.gensat.org*) (Siegert et al., 2009). Transgenic GFP lines have enabled applications such as cell type-specific transcriptome profilings as well as targeted anatomical and physiological analysis (Huang et al., 2003; Siegert et al., 2012). However, functional manipulation of GFP-labeled cell types often requires the use of driver lines such as those that express Cre, which currently exist in limited numbers.

A system converting GFP expression into desired molecular outputs would enable

existing and future transgenic GFP lines to be used directly for gene manipulation of specific cell types. Synthetic RNA devices have been engineered to convert the presence of an intracellular protein into gene expression output (Culler et al., 2010). Although promising, protein-responsive RNA devices await application in animals (Chang et al., 2012). Meanwhile, artificially derived binding proteins, herein including antibodies and unrelated proteins with ideal structures for evolving target recognition (Wurch et al., 2012), are being used intracellularly to target proteins in cells and organisms. Thus far, these reagents are used for target-centric purposes such as protein interference (Jobling et al., 2003), degradation (Caussinus et al., 2012), and modulation (Kirchhofer et al., 2010). Artificially derived binding proteins could possibly be a powerful platform to co-opt intracellular proteins as cell-specific signals that control synthetic circuits, without modifications to the target protein or reliance on the target protein's natural interactions or functions.

We explored whether artificially derived binding proteins can confer GFP with the ability to regulate genes. GFP seems relatively inert in many heterologous systems; it is freely diffusible in the cytoplasm, can enter the nucleus, confers low cytotoxicity and has few interactions with host proteins (Trinkle-Mulcahy et al., 2008). The development of GFP binding proteins (GBPs) from Camelid antibodies (Kirchhofer et al., 2010) has made possible the construction of GBP-fusion proteins non-covalently linking GFP to a variety of proteins in living cells (Caussinus et al., 2012). These reagents, termed nanobodies, are single-chain antigen-binding domains that are relatively small in size (~300-400bp) and can be easily expressed in living cells (Rothbauer et al., 2006). Given the availability of

multiple GBPs, we reasoned that GFP might be used as a scaffold to organize the formation of biologically active complexes. In one scheme, GFP would act like a small molecule "dimerizer", bridging the association of distinct modular domains or protein fragments to reconstitute useful activities such as transcription and recombination (Jullien et al., 2003; Pollock and Clackson, 2002).

Here, we identified pairs of GBPs that can recruit tethered proteins onto the GFP scaffold, providing the means by which GFP-inducible systems can be built. We developed a GFP-dependent transcription system with these reagents, enabling control of any target gene for functional studies across tissues and organisms. The modular design of the transcription system allowed for straightforward and predictable changes to critical features such as DNA binding specificity, transcriptional potency and drug dependency. Our work extends the functionality of GFP into the regulatory realm, thus opening the door to selective manipulation of GFP-labeled cells across transgenic GFP lines and establishing components for the design of synthetic circuits.

Results:

Design and isolation of GFP-dependent transcription factors

In order to use GFP as a dimerizer, one has to identify GBP pairs that can bind to GFP at the same time. Suitable GBP pairs could then bring together fusion protein partners on the GFP scaffold. We obtained six GBPs for this purpose (Kirchhofer et al., 2010). Several GBPs were reported to bind additively to a pre-formed GFP-GBP1 complex when tested as purified proteins *in vitro* (Kirchhofer et al., 2010). However, it was

unclear whether any of the identified pairs could co-occupy GFP, tolerate the addition of fusion partners and induce the formation of biologically active complexes in cells. Furthermore, many possible GBP pair-wise combinations had not been tested for their ability to co-occupy GFP. To address these issues, we performed an *in vitro* reporter screen for GBP pairs that could induce the formation of an active transcription factor (Figure 2.1 and Extended Experimental Procedures in Appendix I). The Gal4 DNA binding domain (DBD) and VP16 activation domain (AD) (Sadowski et al., 1988) were separately fused to GBPs in various configurations and placed under control of the ubiquitous CMV early enhancer/chicken β actin (CAG) promoter (Niwa et al., 1991) (Figure 2.1B). DBD-GBP (DBDG) and AD-GBP (ADG) fusion constructs were screened in pair-wise combinations for GFP-dependent activation of an upstream activating sequence-regulated luciferase (UAS-luc2) reporter in 293T cells. Functional DBDG/ADG pairs will be referred to as transcription devices dependent on GFP (T-DDOG). T-DDOGs employing GBP1+6 or GBP2+7 consistently gave the strongest reporter induction (Figure 2.1C, 2.2 and Appendix I) and became the focus of this study. To specify DBDG+ADG combinations, the DBD-GBPX fusion is listed in regular font, along with the AD-GBPY fusion in superscript, giving DBD-GBPX$^{\text{AD-GBPY}}$. Specific T-DDOG configurations are tabulated in Table S1 of Appendix I.

Characterization of the GFP-dependent transcription system *in vitro*

The induced transcription output in 293T cells was found to be dependent on all components of the system, as removal of GFP, DBDG, or ADG from the transfection

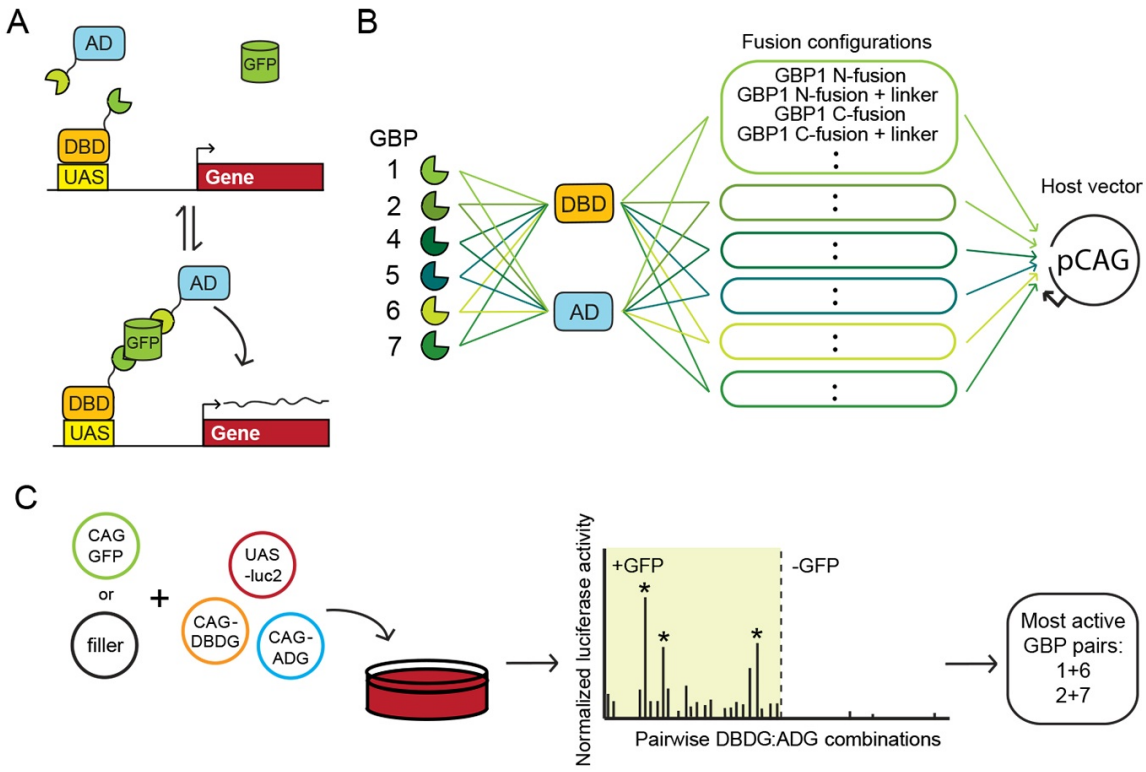


Figure 2.1. In Vitro Screen Used to Identify Functional GBP Pairs for the GFP-Dependent Transcription System. (A) Schematic of GFP-dependent transcription system. DBD, DNA-binding domain; AD, activation domain; UAS, upstream activating sequence. (B) Strategy for making DBD-GBP (DBDG) or AD-GBP (ADG) fusion constructs used in the screen for T-DDOGs. The CAG promoter in pCAG vector was used to drive gene expression. (C) Schematic of typical *in vitro* luciferase screen for functional GBP-fusion combinations capable of inducing GFP-dependent transcription. See also Appendix I.

mixture resulted in loss of reporter activity (Figures 2.2B and Appendix I). Reporter induction was further dependent on the ability of GBP to bind to GFP. Based on the GBP1 + GFP crystal structure (Kirchhofer et al., 2010), we mutated GFP residues expected to directly interact with GBP1. One such variant, GFPmG1, carries the mutations E143K and N147Q. Like GFP, GFPmG1 was localized to the nucleus by the VP16AD-GBP7 fusion protein (Appendix I). However, unlike GFP, GFPmG1 was not localized to the nucleus by the VP16AD-GBP1 fusion protein (Appendix I). In agreement with this, GFPmG1 induced strong UAS-reporter in the presence of Gal4-GBP2$^{VP16\text{-}GBP7}$, but not Gal4-GBP6$^{VP16\text{-}GBP1}$ (Figure 2.2B, 2.2C, and Appendix I). These data confirm a requirement for GFP-GBP interactions, as well as suggest that GBP2 and GBP7 do not depend critically on residue 143 or 147 for binding to GFP.

We also tested whether T-DDOG activity can be controlled by the GFP derivatives cyano and yellow fluorescent proteins (CFP and YFP), and the Discosoma-derived red fluorescent proteins DsRed, mCherry and tdTomato (tdT) (Shaner et al., 2005). CFP and YFP induced Gal4-GBP2$^{VP16\text{-}GBP7}$ activity to a similar extent as GFP (Figure 2.2D). However, CFP had reduced ability to activate Gal4-GBP6$^{VP16\text{-}GBP1}$. This was expected because CFP differs from GFP at the GBP1-interacting residue 147 (Rothbauer et al., 2008). Also as expected, none of the red fluorescent proteins could induce T-DDOG activity. In support of this, red fluorescent proteins were diffusely distributed in the cell even when T-DDOG components were clearly localizing GFP to the nucleus (Figures 2.2C, 2.2E).

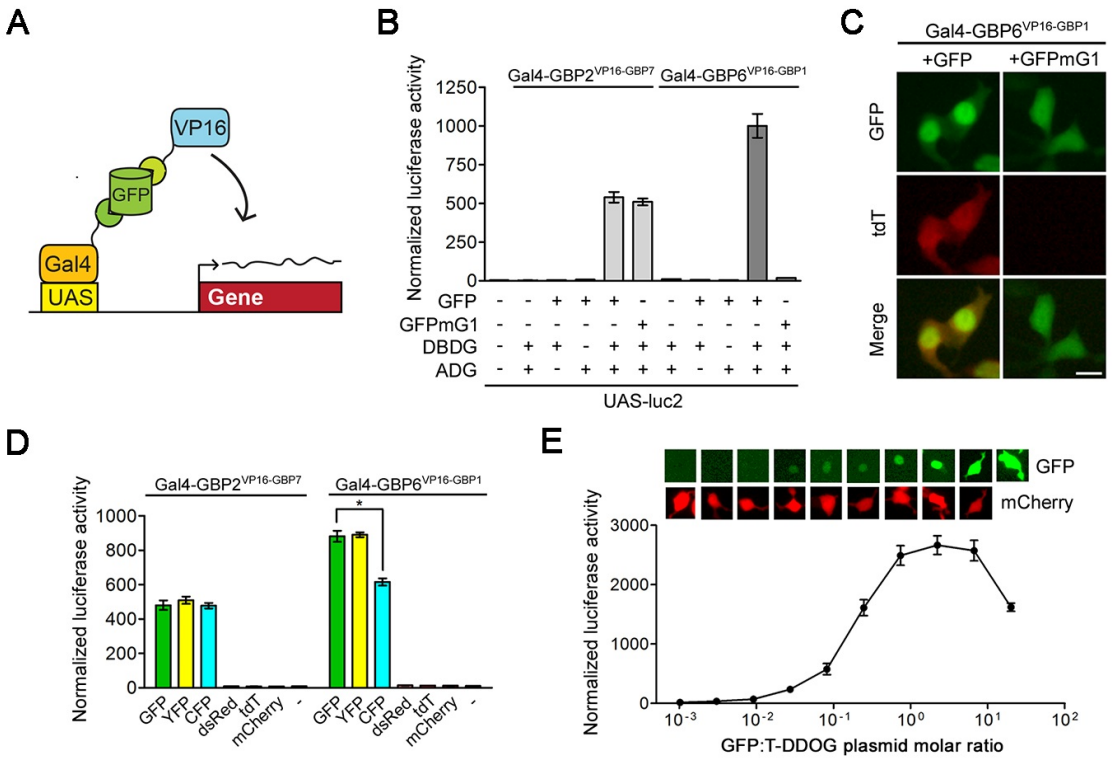


Figure 2.2. Characterization of the GFP-Dependent Transcription System

(A) Schematic of Gal4-based T-DDOGs. (B) GFP-dependent activation of UAS-luc2 by Gal4-GBP6$^{VP16\text{-}GBP1}$ and Gal4-GBP2$^{VP16\text{-}GBP7}$. n = 9. (C) Gal4-GBP6$^{VP16\text{-}GBP1}$ strongly activated UAS-tdT in the presence of GFP. Mutation of GBP1-binding residues in GFP (GFPmG1) abolished tdT activity. Scale bar, 10 μm. (D) Specificity of T-DDOGs for different fluorescent proteins. n = 9; $^{*}p < 0.001$. (E) Activity of Gal4-GBP6$^{VP16\text{-}GBP1}$ in response to a varying amount of transfected GFP plasmids. The transfected DNA amount was kept constant among conditions, with CAG-mCherry (bottom) acting as a filler plasmid to compensate for reduction in GFP (top) plasmids. Panels show representative GFP and mCherry fluorescence in single cells for each corresponding data point below. n = 6. Plots are mean ± SD. See also Appendix I.

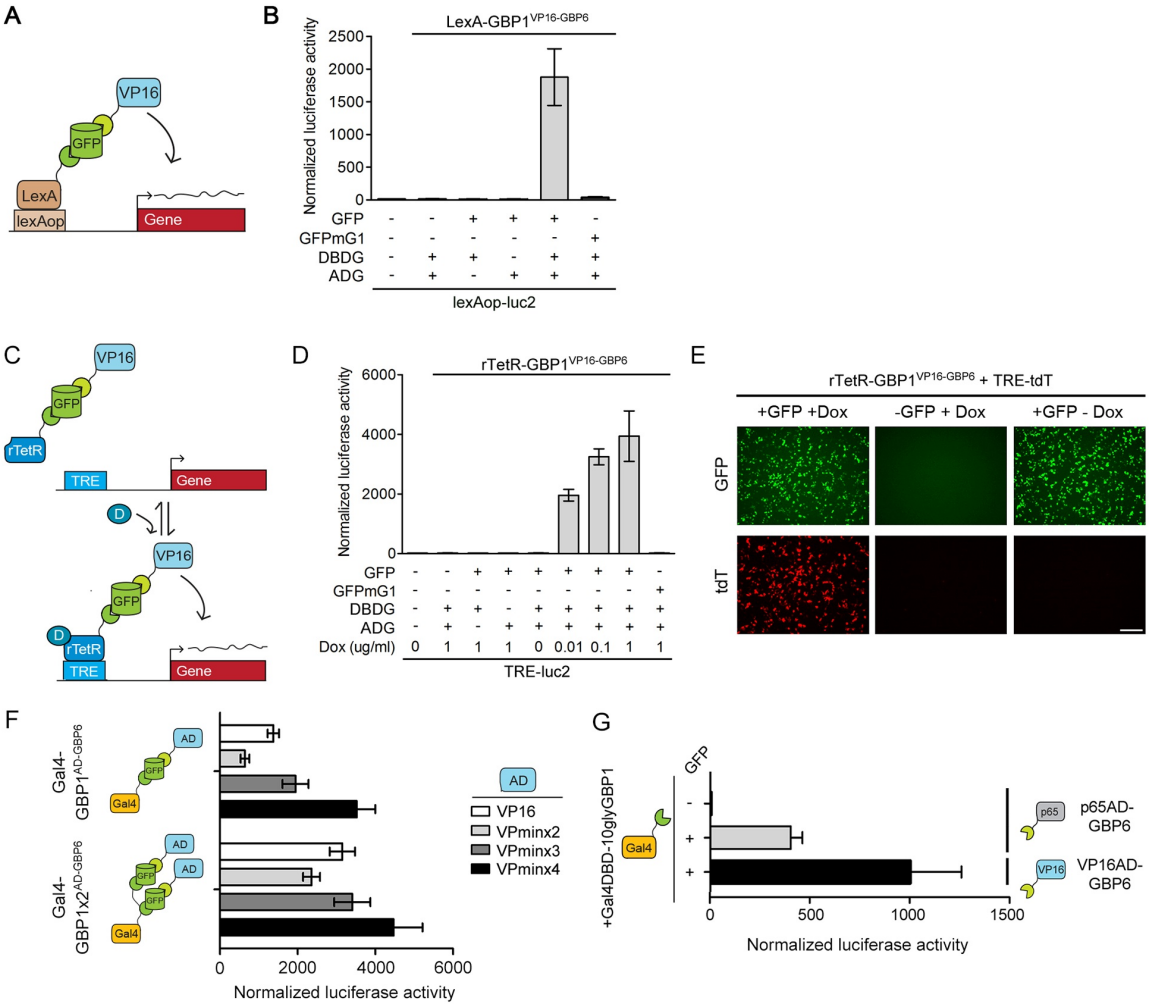
To evaluate the effect of GFP level on T-DDOG activity, we varied the amount of GFP plasmid delivered to 293T cells and examined UAS-luc2 expression in the presence of Gal4-GBP6$^{VP16\text{-}GBP1}$. The observations were consistent with those reported for small molecule dimerizers (Ho et al., 1996). T-DDOG activity increased linearly with the amount of transfected GFP until a certain point, beyond which further increases in GFP led to reduction of activity. The reduced activity is likely due to titration of T-DDOG components by GFP. Interestingly, GFP was highly enriched in the nucleus at levels correlating with the rising phase of the dosage curve, but spread into the cytoplasm at levels correlated with the declining phase of the curve (Figure 2.2E).

Modularity of GFP-dependent transcription system permits various adjustments and fine-tuning

Transcription factors are highly modular (Luan et al., 2006; Sadowski et al., 1988). To exploit this feature for creating a diversity of T-DDOGs with varying properties, we substituted the transcription domains in our original GBP fusion library with other commonly used ones and conducted additional *in vitro* screens. Indeed, we were able expand and diversify the functional repertoire of T-DDOGs. T-DDOGs using the rTetR and LexA DBDs activated reporters bearing their respective binding sequences, tetO (included in tetracycline response element, TRE) and lexAop, only when GFP was present (Figure 2.3A-2.3E) (Butala et al., 2009; Schonig et al., 2010). The activities of rTetR-based T-DDOGs were further found to depend on doxycycline levels (Figure 2.3D). This drug dependency provides temporal control for the system.

Figure 2.3. T-DDOGs Are Highly Adjustable. (A–E) T-DDOGs based on LexA (A) and rTetR (C) DBDs. Doxycycline is "D" in (C). TRE includes seven tetO sequences (C). (B) LexA-GBP1$^{VP16\text{-}GBP6}$ activated a lexAop-luc2 reporter only in the presence of GFP. n = 9. (D) rTetR-GBP1$^{VP16\text{-}GBP6}$ activated TRE-luc2 in a GFP- and doxycycline-dependent manner. n = 6–9. (E) Similar results were seen with TRE-tdT. Doxycycline was used at 1 μg/ml. Images were taken 16 hr post-transfection. (F and G) Tuning T-DDOGs with adjustable DBDs and ADs. (F) Increasing the number of GBP1 on Gal4DBD (n = 6–9) enhanced the transcriptional potency for each ADG (n = 9). (G) Potency of p65AD compared to VP16AD. T-DDOGs used are Gal4-GBP1$^{p65\text{-}GBP6}$ and Gal4-GBP1-B$^{VP16\text{-}GBP6}$. n = 9. Scale bar, 100 μm. Plots are mean ± SD.

Figure 2.3. (Continued)



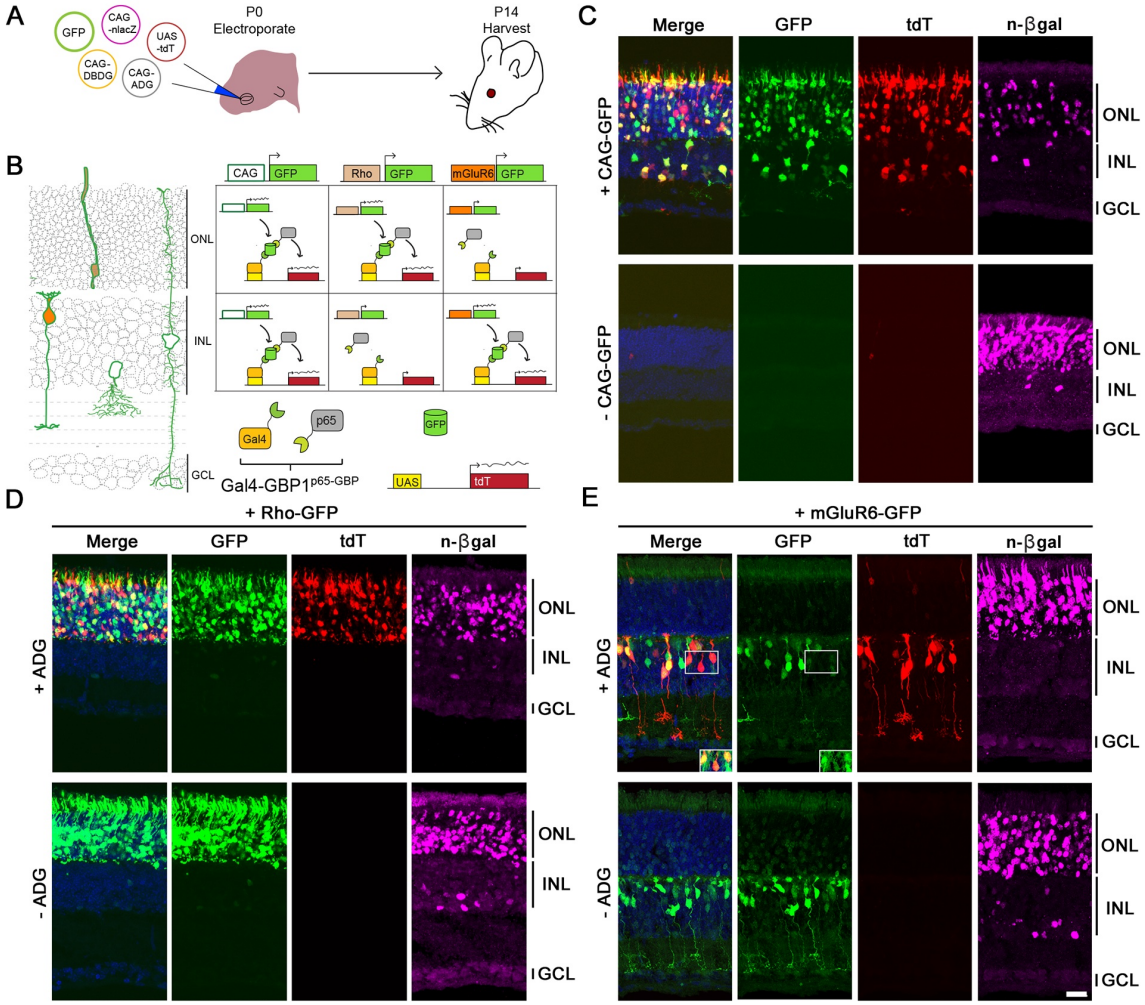
T-DDOGs can also be adjusted to alter their transcriptional potency. The critical region for VP16AD function lies within a 12 amino acid peptide (VPmin) (Baron et al., 1997). We could predictably adjust the transcription activity of Gal4-GBP1$^{AD\text{-}GBP6}$ by either varying the number of VPmin repeats or the number of GBPs fused to the DBD (Figure 2.3F). We further isolated potent T-DDOGs bearing the p65AD (Schmitz and Baeuerle, 1991), an alternative to VP16AD in synthetic transcription systems (Rivera, 1998) (Figure 2.3G). Overall, we consistently isolated potent T-DDOG variants using the GBP1+6 and GBP2+7 combinations, suggesting that these pairs can effectively recruit various combinations of fusion partners onto the GFP scaffold.

The GFP-dependent transcription system can be used in the mouse for cell-specific gene regulation

To evaluate whether GFP can control the activity of T-DDOGs *in vivo*, we used electroporation to introduce GFP, T-DDOGs and UAS-tdT into the murine retina. In our initial tests, we found that overexpression of VP16AD caused mispositioning of rod photoreceptors in the outer nuclear layer (ONL), likely due to squelching of transcription machinery (Appendix I) (Gill and Ptashne, 1988). To address this, we screened devices with the alternative ADs described above for their effects in the retina (Extended Experimental Procedures in Appendix I). We found that T-DDOGs made with VPminx2 and p65 ADs induced little to no disruption of normal rod positioning in the ONL (Appendix I). T-DDOGs bearing p65ADs were used in all subsequent experiments.

Figure 2.4. GFP Controls the Spatial Expression of Genes In vivo. (A) Schematic of experiment. (B and C) (B) (Left) In electroporated retinas, CAG-GFP expresses in multiple cell types (green outline). Rho-GFP expresses in photoreceptors of the ONL (beige fill). mGluR6-GFP expresses in ON bipolar cells of the INL (orange fill). GCL, ganglion cell layer. (Right) Anticipated UAS-tdT expression pattern aligned to left diagram (C) Gal4-GBP1$^{p65\text{-}GBP6}$ induces UAS-tdT only in the presence of GFP. n-βgal (magenta) is an electroporation marker. (D and E) (Top) Rho-GFP and mGluR6-GFP induce tdT expression in rods and ON bipolar cells, respectively. (Bottom) tdT activation depends on ADG. Inset of (E) shows GFP and tdT co-localization upon GFP intensity enhancement. Merge panels includes GFP, tdT, and DAPI channels. Scale bar, 20 μm.

Figure 2.4 (Continued)



We examined how T-DDOG activity would respond to changes in GFP expression in the retina (Figures 2.4 and Appendix I). When GFP was expressed under the broadly active CAG promoter, UAS-tdT was induced in GFP-expressing cell types of both the ONL and inner nuclear layer (INL) (Figure 2.4C). In contrast, little to no tdT signal was detected in electroporated retinas when GFP was excluded (Figures 2.4C and Appendix I). When the GFP expression pattern was manipulated with promoters specifically active in rods (Rho-GFP) (Matsuda and Cepko, 2004) or in ON bipolar cells (mGluR6-GFP) (Kim et al., 2008), the tdT expression pattern shifted accordingly, and was highly restricted to GFP-expressing cells (Figures 2.4D, 2.4E and Appendix I). Cells labeled by the electroporation marker, nuclear β-galactosidase (n-βgal), but not GFP, did not express tdT (Figures 2.4D-2.4E). The efficiency of UAS-tdT activation, adjusting for the probability of a cell receiving all four necessary components for tdT activation, was ~56-93%. Despite the lack of GFP signal amplification with antibodies, more than 90% to 95% of tdT positive cells were positive for GFP expression in all cases (Appendix I). Unexpectedly, we detected faint mGluR6-GFP expression in the ONL, which was not seen without the introduction of T-DDOGs. This low level of GFP induced little to no tdT expression (Figures 2.4E and Appendix I). Follow-up experiments indicated that Gal4-GBP1$^{p65AD-GBP6}$, but not Gal4-GBP2$^{p65-GBP7}$, stabilized a low level of ONL GFP leaking from the mGluR6 promoter (Appendix I). This suggests that Gal4-GBP1$^{p65-GBP6}$ can reveal GFP expression that is normally below the threshold of detection, whereas Gal4-GBP2$^{p65-GBP7}$ allows for gene manipulation without revealing subdetection levels of the native GFP expression pattern. Overall, these results showed that GFP could be used as a cell-specific regulator of T-DDOG activities in the mouse.

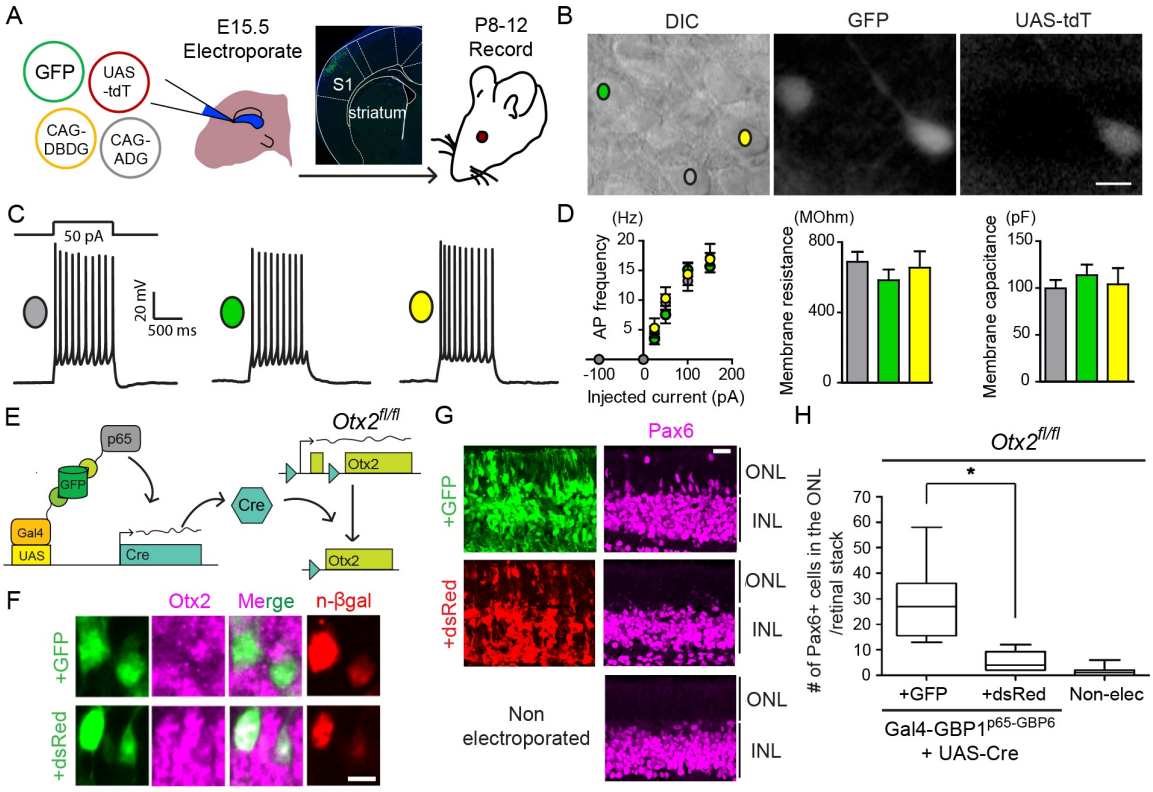
Utility of GFP-dependent transcription system for electrophysiological studies and gene perturbations

To evaluate whether T-DDOGs altered the properties of neurons, we electroporated GFP, T-DDOG s and UAS-tdT into the somatosensory cortex and examined various properties of cortical neurons from ~1.5 week old mouse brains. We compared pyramidal neurons expressing the full set of T-DDOGs, showing both GFP and tdT, with those that expressed GFP alone, as well as with neighboring neurons that lacked fluorescence (Figure 2.5A-2.5D). TdT signal was not observed in GFP-negative neurons in the acute slices (data not shown). We found that excitability and passive membrane properties were similar for the three groups of neurons (Figure 2.5C-2.5D and Appendix I) and were consistent with intrinsic cellular properties previously reported for cortical neurons of this age (Oswald and Reyes, 2008). Moreover, transducing T-DDOGs did not impact morphological features such as dendritic spine density and length (Appendix I). Thus, T-DDOGs are compatible with electrophysiological assays and do not induce functional and structural alterations in the developing brain, within the tested time frame.

We further evaluated the utility of T-DDOGs for deriving biological effects in developing tissues. *Otx2* is a homeobox gene that is necessary for photoreceptor specification in the retina (Nishida et al., 2003). We used GFP to induce Cre-mediated excision of a floxed Otx2 allele ($Otx2^{fl/fl}$) (Tian et al., 2002) in mouse retinas *ex vivo,* with Cre being under the regulation of the UAS promoter (Figure 2.5E-2.5H). This led to the loss of OTX2 protein and the expected ectopic gain of PAX6 in the ONL (Nishida et al., 2003) (Figure 2.5F and 2.5G). Conversely, OTX2 levels were not significantly perturbed when GFP was

Figure 2.5. T-DDOGs Support Electrophysiological and Gene Perturbation Studies in the Central Nervous System. (A) Electroporation setup for neuronal recordings. Micrograph shows GFP in electroporated primary somatosensory cortex (S1). (B–D) (B) Image of an acute brain slice from an electroporated mouse. Scale bar, 10 μm. Three categories of pyramidal layer 2/3 S1 neurons were recorded from brain slices: non-fluorescent controls (gray), GFP+ (green), and GFP+/tdT+ (yellow). (C) Representative single current-clamp trace of action potentials in response to a 50 pA, 1,000-ms-long step current injection. (D) Plots show action potential (AP) frequency upon current injection, as well as input resistance and membrane capacitance of recorded cell classes. $p > 0.5$ for all comparisons (n = 8–10 neurons per condition). Plots show mean ± SEM. (E–H) GFP-dependent excision of $Otx2^{fl/fl}$ in the retina. (E) P0, $Otx2^{fl/fl}$ mouse retina was electroporated with T-DDOG components and UAS-Cre and either CAG-GFP or CAG-DsRed. (F–H) (F) Loss of OTX2 was confirmed by OTX2 immunostaining and (G and H) ectopic PAX6+ ONL cells. n = 10 stacks, 5 retinas per electroporated condition. For non-electroporated retina, n = 19 stacks, 10 retinas. Boxplots show median, maximum, and minimum values. Retinal stacks are 12-μm-thick confocal images. $^{*}p < 0.001$. n-βgal marks electroporated cells in (F). Scale bar, 5 μm in (F) and 20 μm in (G).

Figure 2.5 (Continued)



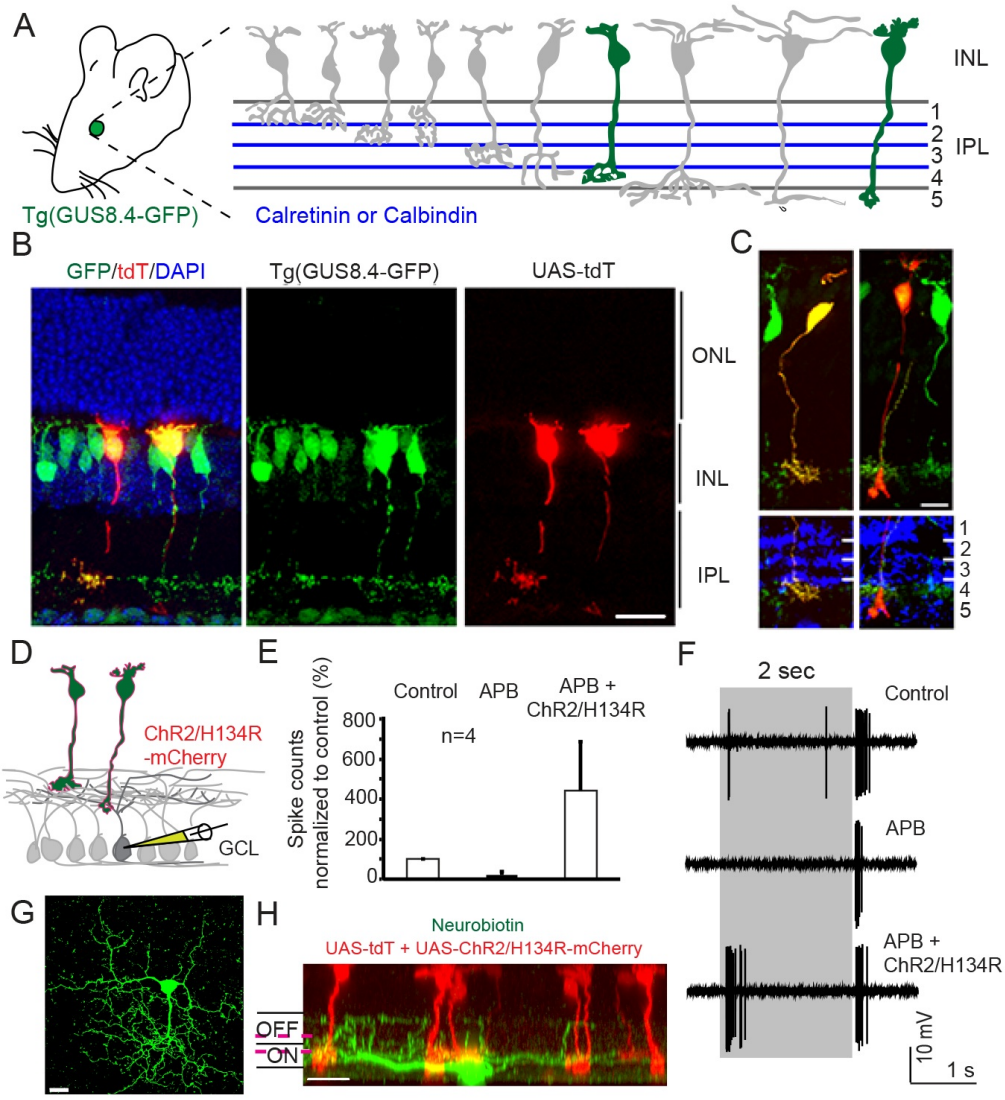
replaced in the same experiment with DsRed (Appendix I). A slight increase in PAX6+ ONL cells above background values was likely due to leakage of UAS-Cre under experimental conditions. Thus, T-DDOGs will be useful for converting GFP expression into desired Cre-mediated genetic changes using a variety of existing conditional alleles. Taken together, these results showed that T-DDOGs are suitable for gene perturbations in the mouse and compatible with assays of cellular function.

Retrofitting transgenic GFP lines for GFP-dependent manipulation of genes and neural circuits

We examined whether T-DDOGs can retrofit transgenic GFP lines for cell-specific gene manipulations. In the mouse retina, visual information detected by rod and cone photoreceptors is transmitted to bipolar cells and ultimately to ganglion cells. Among bipolar cells, the rod bipolar cell type receives input from rods, whereas many types of cone bipolar cells receive input primarily from cones. Currently, almost none of the cone bipolar types can be singly isolated for genetic manipulation, but multiple GFP lines do label subsets of bipolar types (Siegert et al., 2009; Wassle et al., 2009). The α-gustducin-GFP transgenic line, Tg(GUS8.4-GFP) (Huang et al., 2003), expresses GFP in type 7 cone bipolar cells and in rod bipolar cells (Figure 2.6A). Both cell types respond to light increments and are called ON bipolar cells. Introduction of T-DDOGs and UAS-tdT into Tg(GUS8.4-GFP) retinas resulted in tdT induction selectively in these two cell types; identification was based on morphology and axonal stratification in the inner plexiform layer (IPL), aligned to the IPL markers Calbindin or Calretinin (Ghosh et al., 2004) (Figure 2.6A-2.6C). Importantly, 98.9% of tdT+ cells were positive for GFP expression

Figure 2.6. Retrofitting a Transgenic GFP Mouse Line for GFP-Dependent Manipulation of Gene Expression and Neural Circuit Activities. (A) Tg(GUS8.4GFP) expresses GFP in type 7 cone bipolar and rod bipolar cell types (green fill) of the retina. Adopted schematic (Ghosh et al., 2004). (B) Cryosection of electroporated Tg(GUS8.4GFP) retina expressing Gal4-GBP2$^{p65\text{-}GBP7}$ and UAS-tdT. Scale bar, 20 μm. (C) Type 7 (left) and rod bipolar (right) cell types labeled by UAS-tdT. Anti-Calretinin (left) or anti-Calbindin (right) staining identifies specific layers of the IPL. Scale bar, 10 μm. GFP was immunostained in (B and C). (D) Schematic of ChR2 experiment. Electroporated Tg(GUS8.4-GFP) retinas expressing 10×UAS-ChR2/H134R-mCherry and 5×UAS-tdT were analyzed for ChR2-mediated responses in random GCL cells. (E) Cumulative plot of ON responses in GCL cells. Number of spikes counted during the first 300 ms after stimulus onset, normalized to control (minus APB). APB blocks ON responses originating from photoreceptors. Plots are mean ± SEM (n = 4 per condition). (F) Spiking response of a GCL cell. Gray bar, duration of light stimulus. Response to normal light stimuli under control condition (top) or in the presence of APB (middle). Light stimuli focused on INL activate ChR2/H134R in the presence of APB (lower). (G and H) Top and side views of a neurobiotin-filled (green) ganglion cell identified by light stimulation of ChR2. Magenta lines indicate level of anti-Chat bands (not shown). Scale bar, 20 μm.

Figure 2.6 (Continued)



(n=91 cells, sampled from three retinas).

One exciting use of T-DDOGs would be to express light-sensing ion channels in cell types labeled by transgenic GFP for refined, optogenetic probing of neural circuits (Yizhar et al., 2011). We explored this possibility by expressing a UAS-regulated channelrhodopsin-2 (ChR2) variant, H134R (Nagel et al., 2005) in Tg(GUS8.4-GFP)-labeled cells. We asked whether light-driven ChR2 activation in GFP-labeled bipolar cells could trigger downstream spiking responses in cells of the ganglion cell layer (GCL) (Figure 2.6D). Electroporated retinas were presented with two different light stimuli, and recordings were performed on GCL cells. The first stimulus had low light intensity and could evoke photoreceptor-mediated responses in GCL cells but was not bright enough to activate ChR2. We used this stimulus to select GCL cells that responded to both light increments and decrements (ON/OFF cells) (Figure 2.6E-2.6F). We next blocked synaptic communication between photoreceptors and ON bipolar cells with 2-amino-4-phosphonobutyrate (APB) (Slaughter and Miller, 1981) and presented the retina with a brighter light stimulus that could activate ChR2. Because ON/OFF GCL cells receive excitatory input from ON bipolar cells, some of these cells should be connected via excitatory synapses (directly or indirectly) to ChR2-expressing ON bipolar cells. Indeed, the brighter stimulus elicited ON responses in some recorded GCL cells in the presence of APB (Figure 2.6G-2.6H). In contrast, recordings made from ON and ON/OFF GCL cells in non-electroporated regions of multiple retinas did not reveal any response after the onset of the brighter stimulus, in the presence of APB (data not shown). In contrast, recordings made from ON and ON/OFF GCL cells in non-electroporated regions of

multiple retinas did not reveal any response after the onset of the brighter stimulus in the presence of APB (data not shown). Thus, ChR2 activation in rod bipolar or ON cone bipolar cells was robust enough to evoke neurotransmitter release from bipolar cells. Further, the resulting current in GCL cells was large enough to reach spike threshold and evoke spiking responses. These results showed that T-DDOGs could turn on optogenetic tools in transgenic GFP cells, permitting functional interrogation of neural circuits.

Utility of T-DDOGs in zebrafish

In order to determine whether T-DDOGs can direct GFP-dependent activities in other organisms, zebrafish were tested. Here, T-DDOG components were translated from one bicistronic transcript by linking DBDG and ADG components with an internal ribosome entry site (IRES) element. We microinjected RNAs with this structure into ubiquitin-GFP transgenic (Tg(ubi-GFP) (Mosimann et al., 2011) zebrafish embryos in a transient reporter assay. Indeed, mosaic UAS-tdT expression was clearly induced in 78 of 90 injected GFP+ embryos, but not in the 136 injected GFP− embryos (Figure 2.7). This demonstrates the utility of this system across species.

Discussion:

Fluorescent proteins are useful for illuminating cells and cellular processes. Moreover, their apparent lack of connection to many host protein networks makes them ideal scaffolds upon which one can build synthetic complexes with desirable biological activities. We demonstrated this principle here by using GFP to induce formation of a hybrid transcription factor for gene regulation purposes. The ability to use GFP for gene

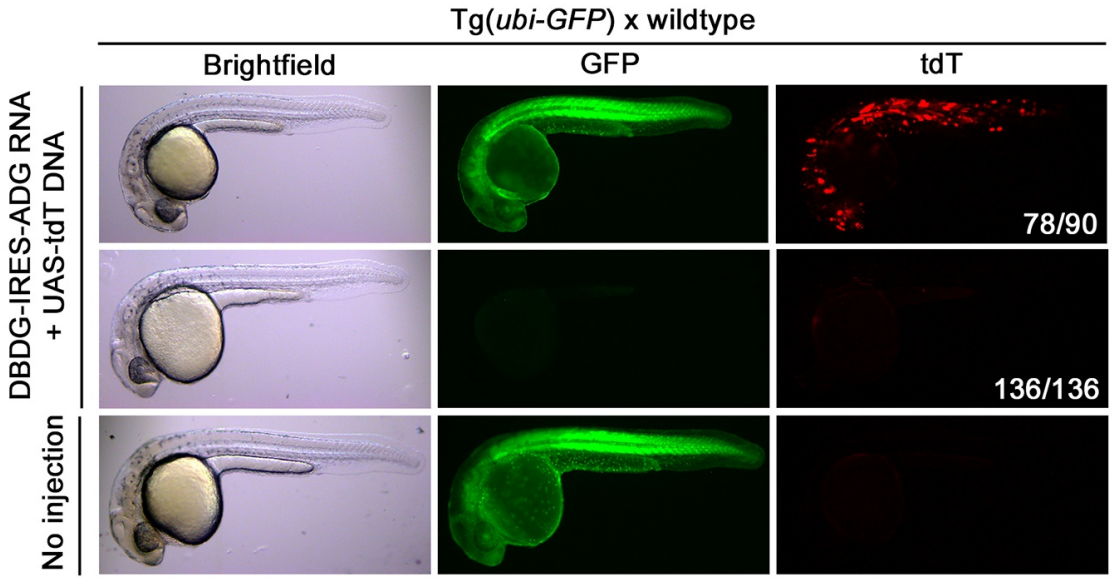


Figure 2.7. GFP-Dependent Transcription in Transgenic Zebrafish Embryos from Tg(ubi-GFP) × wild-type outcrosses were microinjected with DBDG-IRES-ADG (Gal4-GBP1$^{p65\text{-}GBP6}$) RNA and UAS-tdT DNA at the one- to two-cell stage and examined 1 to 2 days post-fertilization. Images represent X number of embryos out of Y number of injected embryos (X/Y), shown in white font in tdT panel.

regulation now enables one to experiment with many GFP-labeled cell types without the need to create new cell-specific driver lines or to discover new cell-specific promoters. This system can be used for gene overexpression and gene deletion (Figure 2.4-2.7) and should be able to perform RNA interference (RNAi) knockdown (Dickins et al., 2007; Dietzl et al., 2007). Activities of the system can be controlled by GFP and its derivatives, but not by red fluorescent proteins, thereby allowing the two types of fluorescent proteins to be used independently in the same experiment. Red fluorescent proteins can likely be used as scaffolds as well. In particular, monomeric variants such as mCherry would be straightforward to use, as they do not undergo obligate dimerization or tetramerization (Campbell et al., 2002).

Selective control of GFP-labeled cells in transgenic GFP organisms

The development and functions of complex multicellular organisms depend upon the activities of a large number of distinct cell types. To investigate these activities in the nervous system, for example, many molecular tools have been developed for anatomical circuit tracing, as well as physiological control (Wickersham and Feinberg, 2012; Yizhar et al., 2011). However, the full potential of these tools can only be realized when one can selectively control them in any cell type in the nervous system. We demonstrated that the diverse transgenic GFP lines available in the mouse and other organisms would be useful for cell-specific manipulation of genes and neural circuits. In the mouse, such manipulations are performed primarily with recombinases such as Cre. Although we anticipate that the number of cell-specific Cre mouse lines will continually increase along with that of GFP lines, each collection of lines will be independently useful for some

applications and will be complementary for other applications. As discussed below, the use of GFP is not necessarily limited to transcriptional control. Nevertheless, gene regulatory systems based on transcription factors differ fundamentally from those based on recombination strategies. As recombination induces alterations to DNA sequences, this typically results in permanent changes in gene expression. In contrast, transcription systems are reversible. Recombination systems are especially useful for targeting cells with a common gene expression history and for long-term transgene expression independent of initial induction signals. However, the irreversibility of recombination events can result in the manipulation of undesired cell types. This should be less of a problem for a transcription system, as continual expression of GFP, in our case, is required to maintain the transcription of target genes. Although GFP may persist for a prolonged period of time after its own transcription has been shut off, this effect may be advantageous in certain applications, such as when it is desirable to achieve a moderately prolonged but reversible gene expression effect. When temporal control of gene expression is desired regardless of GFP expression, rTetR-based T-DDOGs should be useful as they are additionally controlled by drug treatment (Figure 2.3C-2.3E), and could take advantage of the various TRE reagents available (Schonig et al., 2010).

There are additional reasons to use GFP lines for cell-specific targeting. First, not all definable cell types can be specifically targeted by a single driver line (Dymecki et al., 2010). Restriction of target gene expression may be accomplished by intersecting GFP expression with expression of T-DDOG components, other transcription systems, and/or recombination systems. Second, position effects can sometimes unexpectedly activate

GFP in unique cell types. For example, this is thought to be the case for the Tg(GUS8.4-GFP) line and the Tg(Thy1-XFP) collections (Huang et al., 2003) (Feng et al., 2000). Replacement of transgenes typically requires the generation of new transgenic lines and can result in changes to cell specificity of transgene expression. Emerging site-specific, genome-editing strategies hold promise for enabling efficient swapping of transgenes while minimizing changes to cell specificity (Cong et al., 2013; Gohl et al., 2011; Mali et al., 2013). However, it still takes a relatively long time, as well as significant expense, to generate, characterize, and maintain modified transgenic mouse lines.

The GFP-dependent transcription system should find applications beyond mice and zebrafish. As T-DDOGs are built from protein parts commonly used in other model organisms such as Drosophila, other communities can easily adopt T-DDOG components for use in concert with existing GFP driver and responder lines, as well as transient gene delivery vectors. In addition, the modularity of this system allows for a seemingly unlimited number of T-DDOGs to be created according to user demands. Notably, T-DDOGs with customizable DNA-binding specificity (Hsu and Zhang, 2012) would allow for targeted control of endogenous loci without the need for responder cassettes.

Practical considerations for T-DDOG use

Although T-DDOG activities are highly dependent on GFP expression, whether one succeeds in converting an observed GFP expression pattern into corresponding gene output pattern depends on several factors. First, cells expressing GFP at low levels, or transiently, may evade detection in the initial stages of characterization. Such

"background" GFP expression may be detected by certain T-DDOG configurations. Differing fusion protein stability and/or differing GBP affinity for GFP probably contribute to differences in T-DDOG sensitivity. Specifically, we found that Gal4-GBP1$^{p65\text{-}GBP6}$ promoted detection of normally undetectable GFP expression from mGluR-GFP in photoreceptors. However, little to no T-DDOG-mediated expression occurred in these cells. As an mGluR6-driven Cre construct was found to induce recombination in photoreceptors, in addition to bipolar cells (data not shown), we interpret the photoreceptor GFP signal as reflecting leakage from the mGluR6 promoter, as well as stabilization of the leaky GFP by Gal4-GBP1$^{p65\text{-}GBP6}$. Because Gal4-GBP2$^{p65\text{-}GBP7}$ did not reveal the leaky mGluR6-GFP expression and its corresponding T-DDOG did not induce reporter output in photoreceptors, Gal4-GBP2$^{p65\text{-}GBP7}$ may be used for cases when it is not desirable to reveal GFP expression normally below detection threshold. A second issue regards T-DDOG detection of transient GFP expression during early development. This could be addressed by using rTetR-based T-DDOGs for temporal control or by restricting T-DDOGs expression to late progenitors or postmitotic cells, as is possible with electroporation, viral vectors, and/or late-expressing promoters.

Very high levels of GFP expression also require consideration, as too much GFP may saturate GBP binding sites, thereby preventing assembly of T-DDOGs (Figure 2.2E). Nonetheless, we could induce strong T-DDOG readout in Tg(CRX-GFP), a very strong GFP-expressing mouse line (Samson et al., 2009) (data not shown). When necessary, there are approaches to overcome the issue of excessive GFP expression. First, one can capture T-DDOG activity using a recombinase as the T-DDOG readout. We

demonstrated that T-DDOGs could drive expression of Cre to induce irreversible gene expression changes regardless of changes in GFP level. Second, we showed that one could increase the number of GBPs on the DBDG component; this is expected to enhance the GFP binding capacity of the system. Lastly, higher levels of expression of T-DDOGs should be able to balance high GFP levels. Transgenic lines expressing T-DDOGs at high, medium, and low levels should be sufficient for a research community to manipulate a broad range of transgenic GFP lines.

Here, we demonstrated that electroporation or microinjection could immediately be used to deliver T-DDOGs to receptive tissues and organisms for manipulation of GFP-labeled cell types. Additionally, it should require little effort to extend the delivery route to viral vectors. The two components of T-DDOGs can be linked by IRES elements or 2a peptides for expression from a single promoter (Figure 2.7). Each component is relatively small, ranging from ~500 bp to ~1.2 kbp in length, allowing both T-DDOG components to fit into popular viral vectors such as adeno-associated viruses (AAVs) (Yizhar et al., 2011). Responder cassettes can also be delivered virally or by electroporation (Figures 1.4-1.6). Transient delivery methods usually do not provide access to all possible cell types within a tissue or organism; this can be taken into account during experimental design. For example, any retinal cell class is accessible given the right choice of AAV serotypes or electroporation method (Matsuda and Cepko, 2004, 2007; Watanabe et al., 2013). Furthermore, the inherent cell-type specificity of certain gene delivery methods can be exploited to subtract undesired GFP-labeled cell types from being manipulated. For some applications, it will be desirable to deliver T-DDOGs to all cells in a tissue, or

to entire organisms. Transgenic lines expressing T-DDOGs under the control of broadly active promoters would meet these needs, and these tools are under development.

Whenever exogenous components are delivered to cells, one should be cautious of unintended effects. Potential side effects from AD overexpression have long been recognized (Gill and Ptashne, 1988), but applications with transcription factors continue to grow (Schonig et al., 2010; Venken et al., 2011) and drive meaningful discoveries (Kinoshita et al., 2012; Miyamichi et al., 2011). As we found here, problems with an AD may be overcome by testing ADs differing in origin and transcriptional potency. Alternatively, one can reduce expression of a given T-DOGG. IRES-linked cassettes typically give lower expression of genes in the second position relative to those in the first position (Mizuguchi et al., 2000). This could be used to express the DBDG component at high levels while keeping the ADG levels relatively low.

Fluorescent proteins as multi-functional switches for heterologous systems

We believe that this work provides only a glimpse of GFP's regulatory capabilities. To realize GFP's full potential, additional GFP-binding reagents and engineering efforts will be needed to expand its functionality and to improve the performance of GFP-dependent devices. Beyond transcription, the GFP scaffold should be able to regulate other activities such as recombination (Jullien et al., 2003) and proteolysis (Wehr et al., 2006). Although the GBP1 binding epitope on GFP has been revealed by X-ray crystallography (Kirchhofer et al., 2010), it is unclear how GBP2, 6 or 7 bind to GFP. Structural understanding of how GBP pairs co-occupy GFP would facilitate the design of other

GFP-inducible complexes, such as when protein fragments have to be positioned in strict orientations.

Eventually, fluorescent proteins may become preferred transgenes for organisms with long generation times, such as rodents and primates. Because the expression pattern of fluorescent proteins can be characterized from the first set of transgene carriers, any experimental manipulation of labeled cells could be conducted within the same generation by transient device delivery or in the next generation by mating with transgenic device carriers. Also, as one can retroactively build systems to exploit fluorescent proteins for different purposes, it may become unnecessary to generate redundant lines driving different transgenes selectively in the same cell populations. Lastly, engineering of small-molecule ligands that regulate fluorescence would even enable one to use fluorescent proteins exclusively for gene control without interfering with the imaging of other spectrally overlapping probes (Kumagai et al., 2013).

Perspective on targeting intracellular products for cell-specific control

Many intracellular products, such as RNA and proteins, are expressed in a cell-specific manner and could potentially be exploited as spatial signals to control synthetic circuits in multicellular organisms. Here, we demonstrated that artificially derived binding proteins are useful for co-opting an intracellular protein, GFP, for this purpose. Because this approach does not require any modification of the target molecule or rely on the molecule's natural interactions or functions, it may be generalizable to any intracellular product for which artificially derived binding proteins can be selected. Certainly, GFP

seems to be an ideal target because it is an exogenous molecule that shows little connection to host protein networks. However, other exogenous molecules, such as β-galactosidase or Cre recombinase, should also be useful as scaffold proteins upon careful engineering considerations of their unique properties. Furthermore, endogenous molecules probably exhibit a spectrum of connectivity within the host interactome, and a subset might be appropriate for conferring cell-specific manipulations in multicellular organisms. The ability to use intracellular products simply as cell-specific scaffolds would enhance one's ability to target and control cells in nonmodel organisms where transgenic lines are not available.

Experimental procedures

Detailed methods, data analysis, and reagents used can be found in the Extended Experimental Procedures in Appendix I.

Animals - All animal experiments performed were approved by the Institutional Animal Care and Use Committee at Harvard University. Animal information is in the Extended Experimental Procedures in Appendix I.

Molecular Biology and Screens - Using standard techniques, coding sequences of GBPs were fused to those of DBDs or ADs in many configurations, and the products were inserted into pCAG (Niwa et al., 1991). Pairwise combination of DBDG and ADG constructs were then introduced into 293T cells or the mouse retina for screens. See the Extended Experimental Procedures for details.

***In vitro* luciferase assays** - Plasmids encoding CAG-driven GFP, DBDG, and ADG were transfected via polyethyleneimine (PEI) into 293T cells along with plasmids encoding UAS-luc2 and Renilla luciferase or UAS-tdT. Cells were harvested 24 hr later for dual-luciferase assay (Promega) or imaged 16 hr later. All transfections except for the dosage curve were done at a 1:1:1 (GFP:DBDG:ADG) plasmid molar ratio.

***In vivo* electroporations –** P0 CD1 retinas were microinjected with plasmids into the subretinal space and subjected to electroporation (Matsuda and Cepko, 2004). Plasmids encoded DBDG, ADG, CAG-nlacZ (expresses the electroporation marker n-βgal), UAS-tdT, and different promoter-GFP constructs. Electroporated retinas were harvested at P14, immunostained for n-βgal, and imaged by confocal microscopy. At P3–5, Tg(GUS8.4GFP) retinas were electroporated with plasmids encoding T-DDOG components, UAS-driven constructs, and CAG-nlacZ; retinas were harvested between 3 and 4 weeks of age for UAS-tdT detection and between 8 and 10 weeks of age for 10×UAS-ChR2/134R stimulation. Wherever applicable, retinas were immunostained with anti-GFP or anti-DsRed to visualize processes. Anti-Calbindin or anti-Calretinin label layers in the IPL.

Neuronal recordings –C57BL/6 embryos were electroporated with plasmids encoding CAG-driven GFP, DBDG, ADG, and UAS-tdT into the lateral ventricle at embryonic day 15.5. Acute brain slices were prepared from electroporated 1- to 2-week-old mice using standard procedures. Whole-cell current clamp recordings were performed on GFP−, GFP+, and GFP+/tdT+ cortical layer 2/3 pyramidal neurons in regions of dense electroporation. For ChR2/H134R experiment, electroporated retinas from 8- to 10-week-old Tg(GUS8.4GFP) mice were flat mounted, and loose cell-attached patch clamp was

performed on GCL cells that had mCherry/tdT+ bipolar cells in their dendritic fields. Photoreceptors were stimulated by light focused on the outer segments of photoreceptors at a light intensity of 1.3×10^3 R*/s. 20 μM APB was used whenever applicable. ChR2 was stimulated by light focused on the bipolar cell layer at $\sim 10^8$ R*/s for 2 s. Epifluorescence or two-photon microscopes were used to identify fluorescent cells.

Otx2 removal experiment - P0 *Otx2*$^{fl/fl}$ (Tian et al., 2002) retinas were electroporated *ex vivo* (Emerson and Cepko, 2011) with plasmids bearing Gal4-GBP1$^{p65\text{-}GBP6}$, UAS-Cre, and CAG-nlacZ along with either GFP or DsRed-expressing plasmids. Retinas were cultured ex vivo, harvested at P8, immunostained for OTX2, PAX6, and/or n-βgal expression, and subjected to confocal imaging.

Zebrafish microinjections – DBDG and ADG coding sequences of Gal4-GBP1$^{p65\text{-}GBP6}$ were linked by an IRES element, subcloned into pCS2+, and transcribed *in vitro* from the SP6 promoter (mMessage mMachine SP6 RNA Kit, Ambion). RNAs were subjected to LiCl precipitation. One- to two-cell embryos were injected with 40 ng/μl RNA encoding IRES-linked T-DDOG and 25 ng/μl of NotI-linearized UAS-tdT DNA. GFP+ and GFP− embryos, obtained from outcrosses of heterozygote Tg(ubi-GFP) (Mosimann et al., 2011) males to wild-type Tubingen females, were blindly injected with the same RNA/DNA mixture in the same experiment such that the injection success rate for both genotypes should be similar. Injected embryos were incubated at 28°C, and survivors were analyzed for GFP and tdT expression 1 to 2 days post-fertilization.

Statistical Analysis - Two-tailed Student's t test assuming unequal variance was used for all comparisons except cortical recording analysis, for which one-way ANOVA was

used. $p > 0.05$ is judged as statistically significant.

See the Supplemental Information of Appendix I for detailed methods and reagents used.

References

Baron, U., Gossen, M., and Bujard, H. (1997). Tetracycline-controlled transcription in eukaryotes: novel transactivators with graded transactivation potential. Nucleic Acids Res *25*, 2723-2729.

Butala, M., Zgur-Bertok, D., and Busby, S.J. (2009). The bacterial LexA transcriptional repressor. Cell Mol Life Sci *66*, 82-93.

Campbell, R.E., Tour, O., Palmer, A.E., Steinbach, P.A., Baird, G.S., Zacharias, D.A., and Tsien, R.Y. (2002). A monomeric red fluorescent protein. Proc Natl Acad Sci U S A *99*, 7877-7882.

Caussinus, E., Kanca, O., and Affolter, M. (2012). Fluorescent fusion protein knockout mediated by anti-GFP nanobody. Nat Struct Mol Biol *19*, 117-121.

Chalfie, M., Tu, Y., Euskirchen, G., Ward, W.W., and Prasher, D.C. (1994). Green fluorescent protein as a marker for gene expression. Science *263*, 802-805.

Chang, A.L., Wolf, J.J., and Smolke, C.D. (2012). Synthetic RNA switches as a tool for temporal and spatial control over gene expression. Curr Opin Biotechnol.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A.*, et al.* (2013). Multiplex genome engineering using CRISPR/Cas systems. Science *339*, 819-823.

Culler, S.J., Hoff, K.G., and Smolke, C.D. (2010). Reprogramming cellular behavior with RNA controllers responsive to endogenous proteins. Science *330*, 1251-1255.

Dickins, R.A., McJunkin, K., Hernando, E., Premsrirut, P.K., Krizhanovsky, V., Burgess, D.J., Kim, S.Y., Cordon-Cardo, C., Zender, L., Hannon, G.J.*, et al.* (2007). Tissue-specific and reversible RNA interference in transgenic mice. Nat Genet *39*, 914-921.

Dietzl, G., Chen, D., Schnorrer, F., Su, K.C., Barinova, Y., Fellner, M., Gasser, B., Kinsey, K., Oppel, S., Scheiblauer, S.*, et al.* (2007). A genome-wide transgenic RNAi library for conditional gene inactivation in Drosophila. Nature *448*, 151-156.

Dymecki, S.M., Ray, R.S., and Kim, J.C. (2010). Mapping cell fate and function using recombinase-based intersectional strategies. Methods Enzymol *477*, 183-213.

Emerson, M.M., and Cepko, C.L. (2011). Identification of a retina-specific Otx2 enhancer element active in immature developing photoreceptors. Dev Biol *360*, 241-255.

Feng, G., Mellor, R.H., Bernstein, M., Keller-Peck, C., Nguyen, Q.T., Wallace, M., Nerbonne, J.M., Lichtman, J.W., and Sanes, J.R. (2000). Imaging neuronal subsets in transgenic mice expressing multiple spectral variants of GFP. Neuron *28*, 41-51.

Ghosh, K.K., Bujan, S., Haverkamp, S., Feigenspan, A., and Wassle, H. (2004). Types of bipolar cells in the mouse retina. J Comp Neurol *469*, 70-82.

Gill, G., and Ptashne, M. (1988). Negative effect of the transcriptional activator GAL4. Nature *334*, 721-724.

Gohl, D.M., Silies, M.A., Gao, X.J., Bhalerao, S., Luongo, F.J., Lin, C.C., Potter, C.J., and Clandinin, T.R. (2011). A versatile in vivo system for directed dissection of gene expression patterns. Nat Methods *8*, 231-237.

Gong, S., Zheng, C., Doughty, M.L., Losos, K., Didkovsky, N., Schambra, U.B., Nowak, N.J., Joyner, A., Leblanc, G., Hatten, M.E*., et al.* (2003). A gene expression atlas of the central nervous system based on bacterial artificial chromosomes. Nature *425*, 917-925.

Ho, S.N., Biggar, S.R., Spencer, D.M., Schreiber, S.L., and Crabtree, G.R. (1996). Dimeric ligands define a role for transcriptional activation domains in reinitiation. Nature *382*, 822-826.

Hsu, P.D., and Zhang, F. (2012). Dissecting neural function using targeted genome engineering technologies. ACS Chem Neurosci *3*, 603-610.

Huang, L., Max, M., Margolskee, R.F., Su, H., Masland, R.H., and Euler, T. (2003). G protein subunit G gamma 13 is coexpressed with G alpha o, G beta 3, and G beta 4 in retinal ON bipolar cells. J Comp Neurol *455*, 1-10.

Jobling, S.A., Jarman, C., Teh, M.M., Holmberg, N., Blake, C., and Verhoeyen, M.E. (2003). Immunomodulation of enzyme function in plants by single-domain antibody fragments. Nat Biotechnol *21*, 77-80.

Jullien, N., Sampieri, F., Enjalbert, A., and Herman, J.P. (2003). Regulation of Cre recombinase by ligand-induced complementation of inactive fragments. Nucleic Acids Res *31*, e131.

Kim, D.S., Matsuda, T., and Cepko, C.L. (2008). A core paired-type and POU homeodomain-containing transcription factor program drives retinal bipolar cell gene expression. J Neurosci *28*, 7748-7764.

Kinoshita, M., Matsui, R., Kato, S., Hasegawa, T., Kasahara, H., Isa, K., Watakabe, A., Yamamori, T., Nishimura, Y., Alstermark, B*., et al.* (2012). Genetic dissection of the circuit for hand dexterity in primates. Nature *487*, 235-238.

Kirchhofer, A., Helma, J., Schmidthals, K., Frauer, C., Cui, S., Karcher, A., Pellis, M., Muyldermans, S., Casas-Delucchi, C.S., Cardoso, M.C*., et al.* (2010). Modulation of protein properties in living cells using nanobodies. Nat Struct Mol Biol *17*, 133-138.

Kumagai, A., Ando, R., Miyatake, H., Greimel, P., Kobayashi, T., Hirabayashi, Y., Shimogori, T., and Miyawaki, A. (2013). A Bilirubin-Inducible Fluorescent Protein from Eel Muscle. Cell.

Luan, H., Peabody, N.C., Vinson, C.R., and White, B.H. (2006). Refined spatial manipulation of neuronal function by combinatorial restriction of transgene expression. Neuron *52*, 425-436.

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013). RNA-guided human genome engineering via Cas9. Science *339*, 823-826.

Masland, R.H. (2004). Neuronal cell types. Curr Biol *14*, R497-500.

Matsuda, T., and Cepko, C.L. (2004). Electroporation and RNA interference in the rodent retina in vivo and in vitro. Proc Natl Acad Sci U S A *101*, 16-22.

Matsuda, T., and Cepko, C.L. (2007). Controlled expression of transgenes introduced by in vivo electroporation. Proc Natl Acad Sci U S A *104*, 1027-1032.

Miyamichi, K., Amat, F., Moussavi, F., Wang, C., Wickersham, I., Wall, N.R., Taniguchi, H., Tasic, B., Huang, Z.J., He, Z*., et al.* (2011). Cortical representations of olfactory input by trans-synaptic tracing. Nature *472*, 191-196.

Mizuguchi, H., Xu, Z., Ishii-Watabe, A., Uchida, E., and Hayakawa, T. (2000). IRES-dependent second gene expression is significantly lower than cap-dependent first gene expression in a bicistronic vector. Mol Ther *1*, 376-382.

Mosimann, C., Kaufman, C.K., Li, P., Pugach, E.K., Tamplin, O.J., and Zon, L.I. (2011). Ubiquitous transgene expression and Cre-based recombination driven by the ubiquitin promoter in zebrafish. Development *138*, 169-177.

Nagel, G., Brauner, M., Liewald, J.F., Adeishvili, N., Bamberg, E., and Gottschalk, A. (2005). Light activation of channelrhodopsin-2 in excitable cells of Caenorhabditis elegans triggers rapid behavioral responses. Curr Biol *15*, 2279-2284.

Nishida, A., Furukawa, A., Koike, C., Tano, Y., Aizawa, S., Matsuo, I., and Furukawa, T. (2003). Otx2 homeobox gene controls retinal photoreceptor cell fate and pineal gland development. Nat Neurosci *6*, 1255-1263.

Niwa, H., Yamamura, K., and Miyazaki, J. (1991). Efficient selection for high-expression transfectants with a novel eukaryotic vector. Gene *108*, 193-199.

Pollock, R., and Clackson, T. (2002). Dimerizer-regulated gene expression. Curr Opin Biotechnol *13*, 459-467.

Rivera, V.M. (1998). Controlling gene expression using synthetic ligands. Methods *14*, 421-429.

Rothbauer, U., Zolghadr, K., Muyldermans, S., Schepers, A., Cardoso, M.C., and Leonhardt, H. (2008). A versatile nanotrap for biochemical and functional studies with fluorescent fusion proteins. Mol Cell Proteomics *7*, 282-289.

Rothbauer, U., Zolghadr, K., Tillib, S., Nowak, D., Schermelleh, L., Gahl, A., Backmann, N., Conrath, K., Muyldermans, S., Cardoso, M.C*., et al.* (2006). Targeting and tracing antigens in live cells with fluorescent nanobodies. Nat Methods *3*, 887-889.

Sadowski, I., Ma, J., Triezenberg, S., and Ptashne, M. (1988). GAL4-VP16 is an unusually potent transcriptional activator. Nature *335*, 563-564.

Samson, M., Emerson, M.M., and Cepko, C.L. (2009). Robust marking of photoreceptor cells and pinealocytes with several reporters under control of the Crx gene. Dev Dyn *238*, 3218-3225.

Schmitz, M.L., and Baeuerle, P.A. (1991). The p65 subunit is responsible for the strong transcription activating potential of NF-kappa B. EMBO J *10*, 3805-3817.

Schonig, K., Bujard, H., and Gossen, M. (2010). The power of reversibility regulating gene activities via tetracycline-controlled transcription. Methods Enzymol *477*, 429-453.

Shaner, N.C., Steinbach, P.A., and Tsien, R.Y. (2005). A guide to choosing fluorescent proteins. Nat Methods *2*, 905-909.

Siegert, S., Cabuy, E., Scherf, B.G., Kohler, H., Panda, S., Le, Y.Z., Fehling, H.J., Gaidatzis, D., Stadler, M.B., and Roska, B. (2012). Transcriptional code and disease map for adult retinal cell types. Nat Neurosci *15*, 487-495, S481-482.

Siegert, S., Scherf, B.G., Del Punta, K., Didkovsky, N., Heintz, N., and Roska, B. (2009). Genetic address book for retinal cell types. Nat Neurosci *12*, 1197-1204.

Tian, E., Kimura, C., Takeda, N., Aizawa, S., and Matsuo, I. (2002). Otx2 is required to respond to signals from anterior neural ridge for forebrain specification. Dev Biol *242*, 204-223.

Trinkle-Mulcahy, L., Boulon, S., Lam, Y.W., Urcia, R., Boisvert, F.M., Vandermoere, F., Morrice, N.A., Swift, S., Rothbauer, U., Leonhardt, H*., et al.* (2008). Identifying specific protein interaction partners using quantitative mass spectrometry and bead proteomes. J Cell Biol *183*, 223-239.

Tsien, R.Y. (1998). The green fluorescent protein. Annu Rev Biochem *67*, 509-544.

Venken, K.J., Simpson, J.H., and Bellen, H.J. (2011). Genetic manipulation of genes and cells in the nervous system of the fruit fly. Neuron *72*, 202-230Wassle, H., Puller, C., Muller, F., and Haverkamp, S. (2009). Cone contacts, mosaics, and territories of bipolar cells in the mouse retina. J Neurosci *29*, 106-117.

Watanabe, S., Sanuki, R., Ueno, S., Koyasu, T., Hasegawa, T., and Furukawa, T. (2013). Tropisms of AAV for subretinal delivery to the neonatal mouse retina and its application for in vivo rescue of developmental photoreceptor disorders. PLoS One *8*, e54146.

Wehr, M.C., Laage, R., Bolz, U., Fischer, T.M., Grunewald, S., Scheek, S., Bach, A., Nave, K.A., and Rossner, M.J. (2006). Monitoring regulated protein-protein interactions using split TEV. Nat Methods *3*, 985-993.

Wickersham, I.R., and Feinberg, E.H. (2012). New technologies for imaging synaptic partners. Curr Opin Neurobiol *22*, 121-127.

Wurch, T., Pierre, A., and Depil, S. (2012). Novel protein scaffolds as emerging therapeutic proteins: from discovery to clinical proof-of-concept. Trends Biotechnol.

Yizhar, O., Fenno, L.E., Davidson, T.J., Mogri, M., and Deisseroth, K. (2011). Optogenetics in neural systems. Neuron *71*, 9-34.

Chapter Three

A GFP-Dependent Cre Recombinase for Retrofitting Transgenic GFP Lines

A GFP-Dependent Cre Recombinase for Retrofitting Transgenic GFP Lines

Authors: Chung Yiu Jonathan Tang[1,2], Sylvain Lapan[1,2], Eugene Drokhlyansky[1,2] and Constance L. Cepko[1,2]

[1]Howard Hughes Medical Institute

[2]Departments of Genetics and Ophthalmology

Harvard Medical School, Boston, MA 02115, USA

This chapter contains the manuscript titled "A GFP-Dependent Cre Recombinase for Retrofitting Transgenic GFP Lines" in preparation. It is modified to better fit the style of this dissertation. Jonathan C.Y. Tang is first author on this manuscript. J.C.Y.T initiated the project. J.C.Y.T. conducted experiments and analysis, S.L. and E.D. constructed the ChAG-loxP-BFP-loxP-mCherry construct. S.L. electroporated CRE-DOG constructs into the Prox1-GFP line. C.L.C. supervised the project. J.C.Y.T and C.L.C wrote the paper.

Summary

The expansion of GFP-regulated activities would enhance the utility of GFP for gene manipulation. Here, we developed a Cre recombinase dependent on GFP (CRE-DOG). We systematically optimized the efficiency and GFP specificity of CRE-DOG. In the mouse, CRE-DOG induces loxP recombination in a GFP-specific manner for activation of loxP-based responder cassettes. CRE-DOG can be used to retrofit transgenic GFP lines as tools for cell-specific gene manipulation in the mouse nervous system and likely beyond.

Introduction

Green Fluorescent Protein (GFP) from *Aequorea victoria* has become an indispensible tool in biology (Chalfie, 2009; Shimomura et al., 1962; Tsien, 1998). GFP's ability to fluoresce without cofactors has enabled its use as a marker of gene expression, protein localization, and as a reporter physiological signals such as calcium (Chalfie et al., 1994; Nakai et al., 2001; Tsien, 1998). Indeed, much investment has been made towards the generation of thousands of transgenic GFP lines using GFP as a marker of gene expression (Chalfie, 2009). This is particularly significant in the mouse, where a systematic effort to generate transgenic GFP lines has resulted in many lines that label unique cell populations in complex tissues, such as the central nervous system (gensat.org) (Gong et al., 2003; Heintz, 2004; Siegert et al., 2009). Unlike other common molecular tools such as transcription factors or site-specific recombinase, GFP does not have any known regulatory function, and it is thus limited to imaging studies. To take advantage of existing transgenic GFP lines for cell-specific manipulations, I developed a

GFP-dependent transcription system as described in Chapter 2. While such a system has opened the door for gene manipulation of GFP-labeled cells, it suffers from potential toxic effects caused by overexpression of the activation domain (Tang et al., 2013). Also, although a transcription system is reversible and has its advantages, it makes it less ideal for lineage tracing, which actually prefers irreversible capturing of molecular events. Furthermore, many transgenic mouse lines and viral gene delivery tools have been generated and designed to respond to Cre (gensat.org) (Yizhar et al., 2011). Thus, a GFP-dependent Cre recombinase would simplify the means by which one can access these tools for experimentation.

Here, I describe a simpler strategy for converting GFP expression into Cre recombinase activity. The system uses GFP as a dimerizer to bring together split Cre fragments (Jullien et al., 2003) for recombination of loxP-based responder cassettes. The GFP-dependent Cre recombinase has been optimized to minimize background recombination; the tightness of this system is demonstrated in the mouse retina by the retrofitting of transgenic GFP lines to target specific cell populations.

Results:

Isolation of a GFP-dependent Cre recombinase

In Chapter 2, we identified two GBP pairwise combinations, GBP1+GBP6 and GBP2+GBP7, which could be used to assemble a hybrid transcription factor on the GFP scaffold in living cells (Tang et al., 2013). These GBP pairs might be able to reconstitute split-protein fragments for recombination and proteolysis. We focused on isolating a

GFP-dependent Cre recombinase (Figure 3.1A), given its potential use for taking advantage of loxP-based conditional alleles and responder cassettes built for the mouse and other model organisms. We began by screening fusion constructs bearing GBP and split Cre components for GFP-dependent recombination of a CAG-loxP-Neo-loxP-luciferase reporter (CALNL-luc2) reporter (Figure 3.1B) (Jullien et al., 2003). Although we found GFP-dependent Cre-recombination of reporters, the recombination efficiency was unsatisfactorily low, and high background was often observed (data not shown). The low efficiency could be due to the need to form a ternary complex composed of GFP and split-Cre components in highly specific orientation. The efficiency might be increased if GFP instead induced formation of full-length Cre recombinase via a protein splicing mechanism (Vila-Perello and Muir, 2010). We thus inserted the artificially split S. cerevisiae vacuolar ATPase (VMA) intein elements (Tyszkiewicz and Muir, 2008) into the GBP-split Cre constructs. We term these fusion constructs N-terminal Cre-intein-GBP (N-CreintG) or C-terminal Cre-intein-GBP constructs (C-CreintG). In a reporter assay screen, we found that most N- and C-CreintG pairwise combinations again suffered from a low GFP-dependent recombination efficiency and high background, but a small subset of constructs based on the GBP2+GBP7 combination gave ~20 to 60-fold induction of recombination in a GFP-dependent manner, along with reasonably low background in the absence of GFP, at 3 to 4-fold above control. Surprisingly, the GBP1+GBP6 pair did not yield any high efficiency recombination. This suggests that the GBP1+GBP6 pair does not orient the tested split Cre fragments in a manner suitable for recombinase reconstitution. We collectively termed the most successful pairs Cre-recombinase Dependent on GFP (CRE-DOG) (Figure 3.1A).

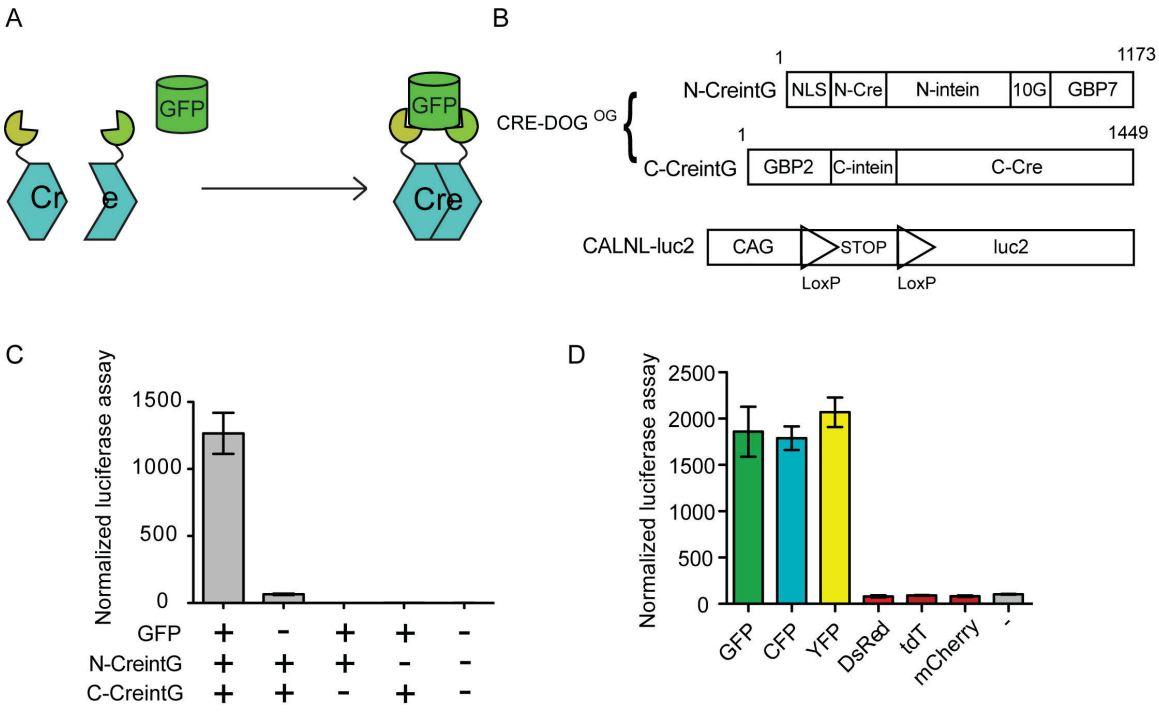


Figure 3.1. A GFP-dependent Cre recombinase. (A) Schematic of GFP-dependent recombinase system. GFP brings together split Cre fragments via interaction with GBPs (pacman shapes). (B) Components used to test CRE-DOG activity. N-CreintG and C-CreintG make up CRE-DOG^{OG}. Triangle shapes in CALNL-luc2 correspond to LoxP sequences. CAG-loxP-Neo-loxP-luc2 floxed reporter (CALNL-luc2) is used to assay recombination (C) GFP-dependent recombination in 293T cells transfected with CRE-DOG^{OG}, CALNL-luc2 and Renilla luciferase plasmids. n=9 for all conditions. (D) CRE-DOG^{OG} is active with GFP derivatives cyano fluorescent protein (CFP) and yellow fluorescent protein (YFP), but not the Discosoma-derived DsRed, tdTomato (tdT), and mCherry. n=9 for all conditions. Cells in (C-D) were harvested 24 hour post-transfection. Normalization is against Renilla luciferase activity. Plots are mean +/- standard deviation.

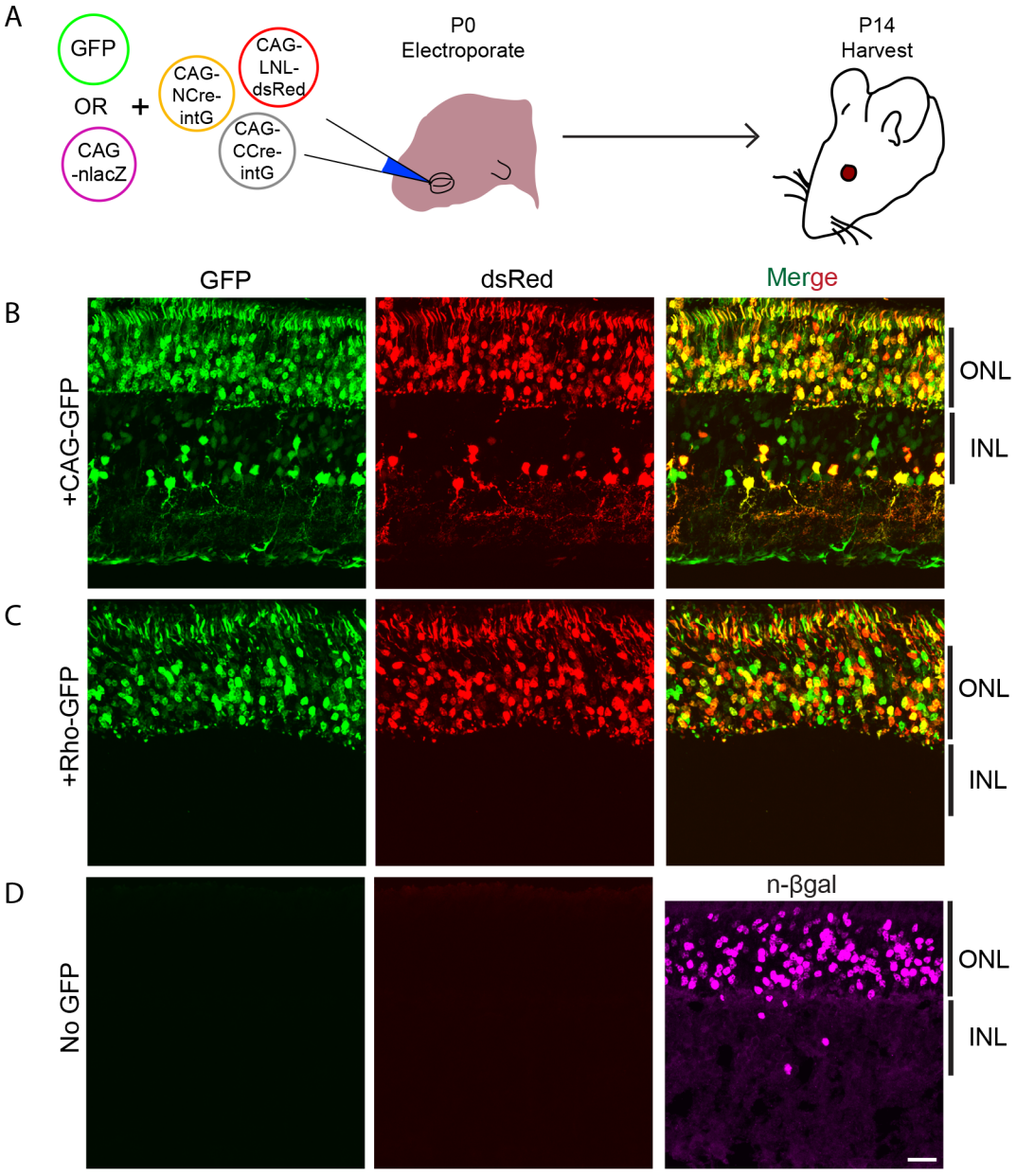
Characterization of the GFP-dependent Cre recombinase

We selected the most potent and GFP-dependent fusion pair, or the Original CRE-DOG (CRE-DOG^{OG}), for further analysis. CRE-DOG^{OG} refers to the combined action of two fusion protein components, N-CreintG and C-CreintG (Figure 3.1B). Using CALNL-luc2 to quantify CRE-DOG activity, we found that recombination only occurred in the presence of all components, but not when any component was removed from the transfection mixture (Figure 3.1C). Consistent with previous observations that GBP2+GBP7 based T-DDOGs can facilitate transcription in response to the GFP derivatives cyano (CFP) and yellow (YFP) fluorescent proteins, CRE-DOG^{OG} also induce recombination in the presence of CFP and YFP (Figure 3.1D). In addition, CRE-DOG^{OG} activity did not respond to DsRed or its derivatives, mCherry and tdTomato (Shaner et al., 2005) (Figure 3.1D).

To assess whether CRE-DOG^{OG} will be useful for *in vivo* studies, we examined its ability to induce GFP-dependent recombination in the mouse retina (Figure 3.2 and 3.3). We electroporated plasmids encoding CAG-GFP or CAG-nlacZ along with CRE-DOG^{OG} and a floxed-DsRed reporter into P0 mouse retinas (Figure 3.2A). At P14, we found that CAG-GFP electroporated retinas showed strong activation of the floxed reporter (Figure 3.2B). In contrast, retinas electroporated without GFP plasmids, but expresses the electroporation marker n-βgal, failed to activate DsRed expression (Figure 3.2D). Further, CRE-DOG can act in a cell-specific manner, as the activation of DsRed can be restricted to rod photoreceptors when Rho-GFP is used in place of CAG-GFP (Figure 3.2C). In contrast to the T-DDOG system, which carries the caveat of being potentially

Figure 3.2. GFP directly controls Cre recombination in a cell-specific manner *in vivo.* (A) Schematic of experiment. CAG-NCreintG and CAG-CCreintG together express CRE-DOG^{OG} . CAG-LNL-DsRed is a Cre-dependent reporter. CAG-nlacZ serves as an substitution plasmid for GFP. (B-D) In electroporated retinas, CRE-DOG^{OG} induces recombination and expression of CALNL-DsRed reporter in a GFP-dependent manner. As seen by comparing the broadly expressing CAG-GFP (A) and the rod-specific Rho-GFP (B), the spatial expression pattern of DsRed is changes according to the GFP expression pattern. CAG-nlacZ (which gives the n-βgal product) cannot substitute for CAG-GFP to express DsRed. (D). Scale bar, 20 μm. See Figure 3.3. for quantifications and sample size.

Figure 3.2 (Continued)



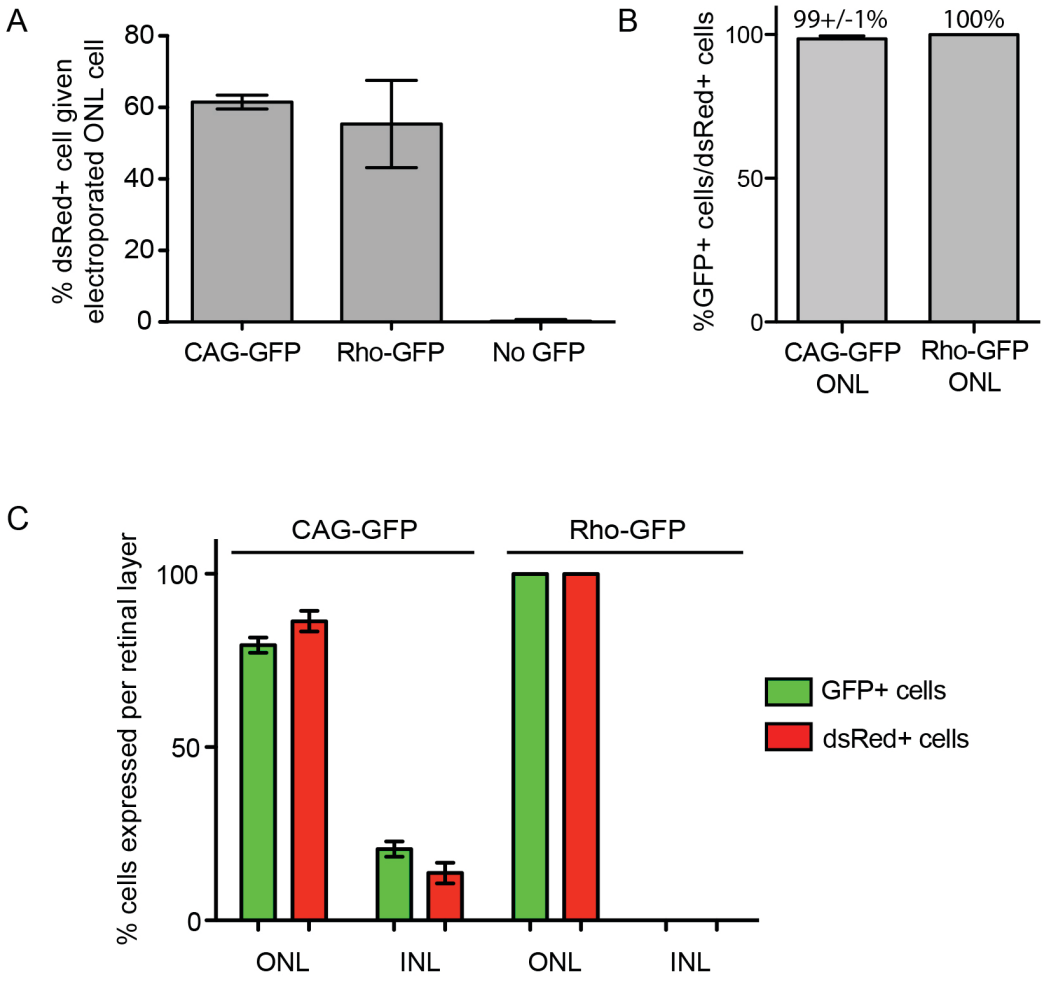


Figure 3.3. Quantification summary of data from Figure 3.2. (A) Efficiency of CRE-DOG^{OG} recombination in electroporated retinas. An electroporated cell is defined by GFP expression in CAG-GFP and Rho-GFP conditions, and by nβgal expression in No GFP condition. Counted cell types are rod photoreceptors in the outer nuclear layer (ONL) (B) GFP specificity of floxed DsRed+ ONL cells. (C) Expression pattern of floxed DsRed+ cells compared to GFP+ cells in ONL versus inner nuclear layer (INL). Sample size (retinas): CAG-GFP, n=3; Rho-GFP, n=4; No GFP, n=3. Consistent results were obtained in an independent experiment.

toxic to cells as a consequence of the squelching phenomenon caused by activation domains (Gill and Ptashne, 1988), CRE-DOG^{OG} did not induce any noticeable abnormality in retinal cell phenotypes (data not shown).

Factors affecting CRE-DOG activity and specificity

Although inclusion of the intein sequences were intended to create a Cre protein via protein splicing, it was possible that CRE-DOG^{OG} activity was due to the mere formation of a GFP-dependent complex, rather than, or in addition to, protein splicing. To determine whether protein splicing was required for CRE-DOG^{OG} activity, we analyzed CRE-DOG^{OG} mutants for recombinase activity. We first introduced mutations known to abolish intein splicing activity. The split-intein of VMA1 requires a cysteine residue at both boundaries between the N-intein and C-intein with the fusion protein partners. Substitution of both residues with alanine or serine either abolish or reduce protein splicing (Anraku et al., 2005). Thus, we compared CRE-DOG^{OG} activity with double alanine and double serine substitutions at the cysteine positions. GFP-dependent recombination was not or only slightly affected with mutant CRE-DOG^{OG}, suggesting that protein splicing does not play a significant role in the action of CRE-DOG^{OG} (Figure 3.4A). Second, we found that the split VMA element on each components of CRE-DOG^{OG} independently promoted GFP-dependency of recombination. Inclusion of one of either VMA element in the fusion constructs was sufficient to confer improved GFP specificity of recombination events compared to the situation where both VMA elements were removed from the fusion components (Figure 3.4B). Since the influence of each split VMA elements seems to be independent of the presence of their protein splicing

partners, the VMA elements might contribute to the GFP specificity of CRE-DOG^{OG} in a protein splicing independent manner. To more directly determine whether split Cre fragments are spliced together, we probed for full-length Cre formation in the presence or absence of GFP. We used the FLAG epitope tag to label the C-CreintG component, reasoning that protein splicing of the split Cre fragments should result in a shift in size from ~52 kDa to ~37 kDa (Figure 3.4C). We tested and confirmed that CRE-DOG^{OG} made of FLAG-tagged C-CreintG (CRE-DOG^{OG}-FLAG) was about as active in a floxed luciferase assay as untagged CRE-DOG^{OG} (data not shown). However, under conditions in which GFP could induce CRE-DOG^{OG}-FLAG activity, we failed to observed the appearance of the expected 37kDa band in the presence of GFP, suggesting against protein splicing in mediating GFP-dependent recombination (Figure 3.3D). Thus, we discovered an unexpected, beneficial effect of split VMA intein elements on the GFP-specificity of GBP-split Cre fragments. As these combined fusion proteins fulfill the functional requirement of a GFP-dependent Cre recombinase, we continued to optimize these fusion constructs and tested them for applications.

Optimizing and modifying CRE-DOG activity

We took steps to optimize CRE-DOG^{OG} activity by looking for variants that increased GFP-dependent recombination and/or reduced GFP-independent recombination. We first tested whether the VMA portion of CRE-DOG^{OG} could be truncated without compromising its beneficial effects on GFP specificity. We arbitrarily divided the 184 aa N-terminal portion of VMA (N-VMA), as found in N-CreintG, into four segments for truncation studies (Figure 3.5A). The majority of the N-VMA sequence could actually be

Figure 3.4. CRE-DOG appears to act independently of protein splicing. (A-B) 293T cells transfected with floxed luciferase reporter, GFP or filler plasmids, and CRE-DOG^{OG}-encoding plasmids were harvested 1 day post-transfection for analysis. (A) Mutations known to disrupt protein splicing ability of VMA intein did not cause significant alteration of CRE-DOG^{OG} activity. N-splicing site refer to a cysteine residue in the N-intein portions of N-CreintG of CRE-DOG^{OG}. C-splicing site refer to a cysteine residue in the C-intein portions of C-CreintG of CRE-DOG^{OG}. (B) Both intein portions were required for the tight GFP-dependent recombination activity of CRE-DOG^{OG}. Int in the name indicates presence of intein portion. Removal of either intein portion had adverse effect on CRE-DOG^{OG} specificity for GFP, but the result also indicates that either portion conferred GFP-specificity on the system. (C) Testing for protein splicing mechanism for CRE-DOG^{OG} function. FLAG tag is used to label the C-terminal component of CRE-DOG^{OG} (~52 kDa). If protein splicing occurs to generate full-length Cre recombinase, then a ~37 kDa protein product should appear on western blot probing for FLAG epitope. (D) Result of western blot testing the protein splicing model in (C). 293T cells transfected with the indicated components (CRE-DOG^{OG}-FLAG and GFP) were lysed 1 day post-transfection. Lysate supernatant were probed for protein products labeled by the FLAG tag. Results failed to support protein splicing mechanism with absence of ~37 kDa band. Consistent results were obtained in an independent experiment with whole cell lysate loading for western blotting.

Figure 3.4 (Continued)

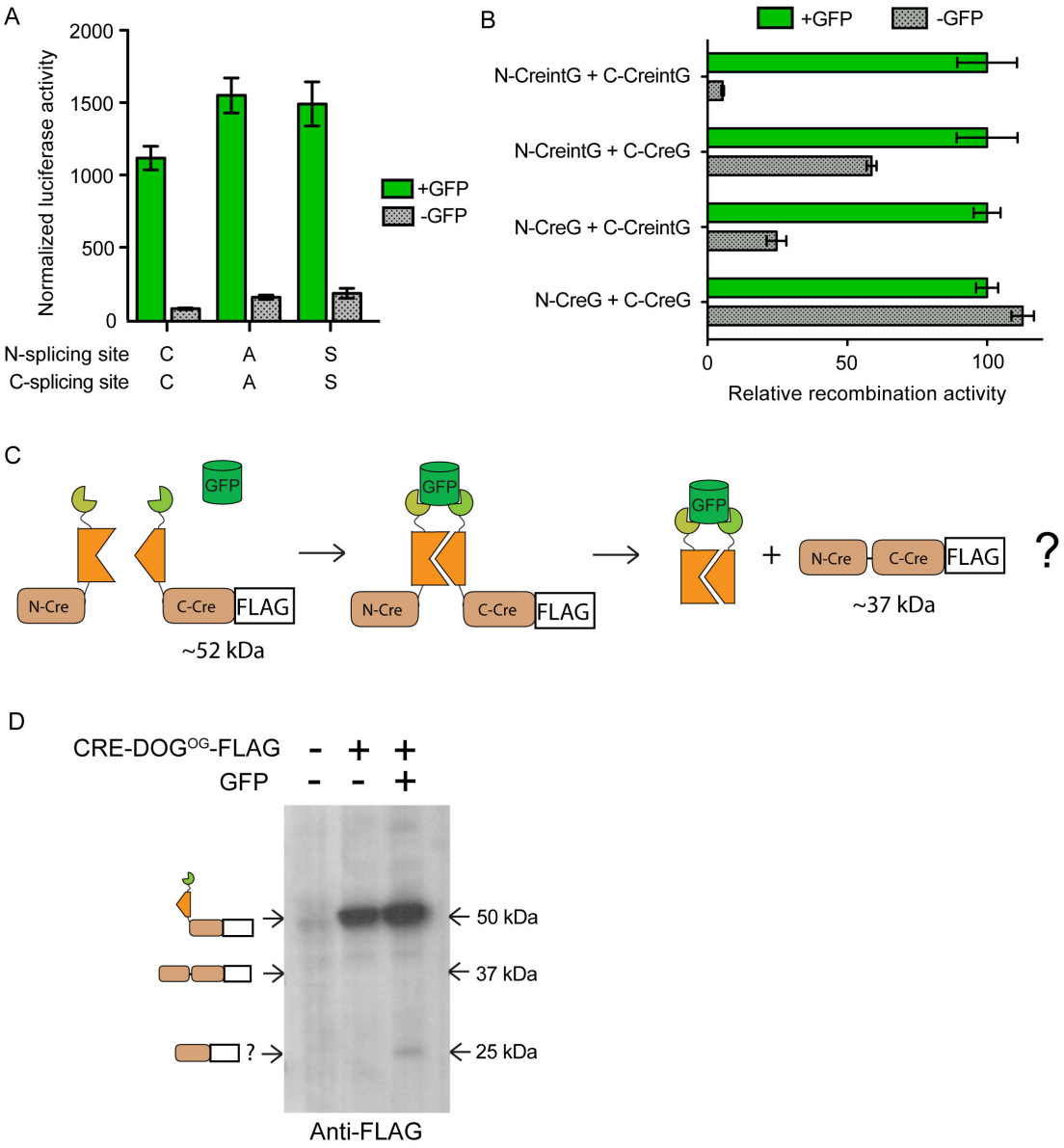
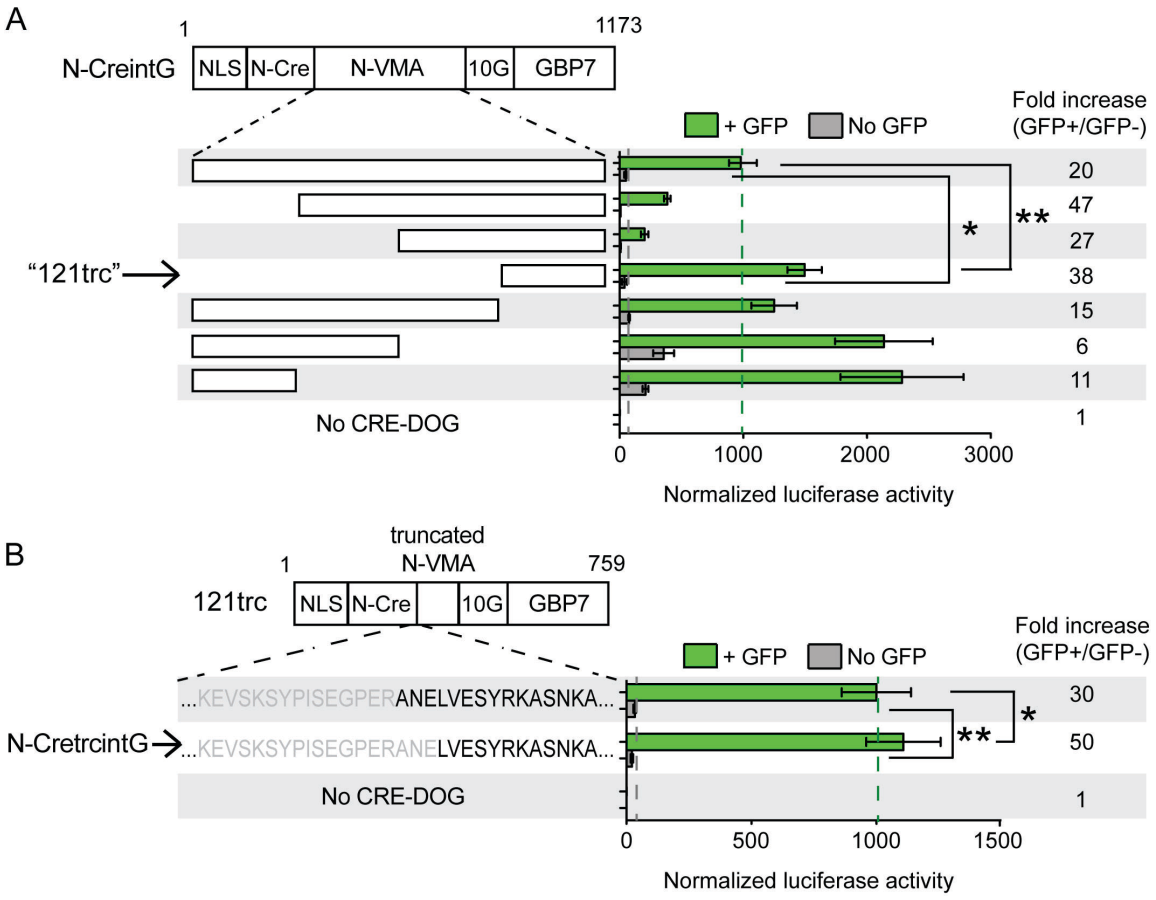


Figure 3.5. Truncation of N-VMA intein fragment promotes enhanced GFP-specificity and activity of CRE-DOG. (A-B) Normalized reporter luciferase assay done in 293T cells and harvested 24 hours post-transfection. Truncation split Cre constructs were adjusted to be equimolar in transfected plasmid amount. **(**A) Broad truncation scans along N-VMA intein fragment reveals a 43 amino acid fragment that promotes increased GFP-dependent recombination and specificity in the context of the split-Cre/GBP7 fusion construct. This truncated construct is code named "121trc". n= 15-18 per condition, pooled from 2 independent experiments. *, $p<0.01$. **, $p<0.0001$. (B) Finer resolution truncation scans along the 43 amino acid N-VMA fragment reveals that an additional truncation of 3 amino acid increased GFP-dependent recombination and specificity in the context of the 121trc construct. This construct is named N-CretrcintG. n= 19-21 per condition, pooled from 3 independent experiments. *, $p<0.05$. **, $p<0.0001$.

Figure 3.5 (Continued)

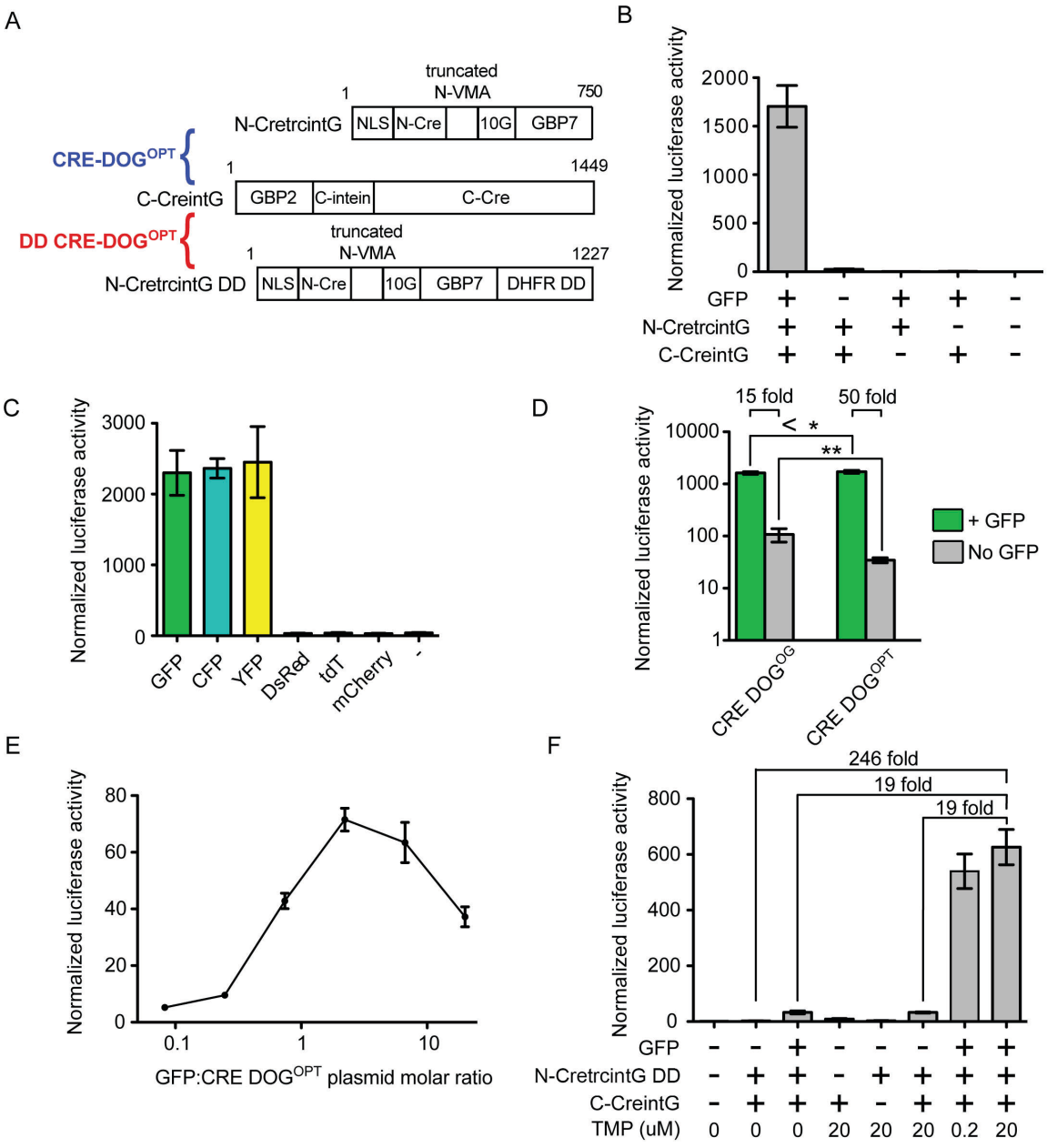


removed without compromising GFP specificity (Figure 3.5A). The smallest essential segment for conferring GFP-dependent recombination is a 46 aa fragment in the C-terminal portion of N-VMA (N-VMAC 139-184). Next, we reintroduced portions of the removed N-VMA residues to N-VMAC 139-184, or deleted additional residues from N-VMAC 139-184. We first reintroduced or truncated from the N-terminal end of N-VMAC 139-184 at 3 amino acid resolution and found that addition or subtraction of 3 amino acids led to further decrease in GFP-independent recombination without compromising GFP-dependent recombination (data not shown). We next scanned for changes by adding or deleting N-VMA residues one amino acid at a time, but found no further improvements within a 10 amino acid window (data not shown). Thus, these screens resulted in isolation of a truncated fusion protein, hereafter named N-CretrcintG, which when combined with C-CreintG, gives the most GFP-specific recombination activity (Figure 3.5B).

We further characterized the N-CretrcintG and C-CreintG combination, hereafter referred to as CRE-DOG^{OPT}, for its activity in several experimental settings and applications (Figure 3.6A). CRE-DOG^{OPT} activity was found to depend on all components of the system, was specific for GFP and its derivatives, and could induce recombination of floxed or FLEX reporter constructs (Figure 3.6B-C) (data not shown for FLEX). In comparison to CRE-DOG^{OG}, CRE-DOG^{OPT} gives significantly higher GFP-dependent recombination as well as significantly lower GFP-independent background activity (Figure 3.6D). This results in an overall 3-4 times higher GFP-induced fold increase in GFP-dependent activity over CRE-DOG^{OG} (Figure 3.6D). Further as expected, we

Figure 3.6 Characterization of the optimized CRE-DOG. (A) Components of CRE-DOG^{OPT} and DD-CRE-DOG^{OPT}. DD, destabilizing domain. NLS, nuclear localization signal. (B) GFP-dependent Cre recombination in 293T cells transfects with CRE-DOG^{OPT} components along with CAG-GFP or filler plasmids. n=18 per condition. (C) CRE-DOG^{OPT} is function with the GFP derivatives CFP and YFP, but not the red fluorescent proteins DsRed, tdT and mCherry. n=18 per condition. (D) CRE-DOG^{OPT} has greater activity and lower background than CRE-DOG^{OG}. < * indicate CRE-DOG^{OPT} has significantly greater activity than CRE-DOG^{OG}. n=12 per condition. * indicates $p<0.05$. ** indicates $p<0.0001$. Two-tailed Student's t test assuming unequal variance. (E) GFP dosage dependency of CRE-DOG^{OPT}. n=6 or 9 per condition. (F) GFP and TMP dependency of DD-CRE-DOG^{OPT} in transfected 293T cells. n=12 per condition. All cells were harvested 1 day post-transfection. Plots are mean +/- standard deviation.

Figure 3.6 (Continued)



observed a dosage dependent effect of GFP on CRE-DOG^{OPT} activity. Increasing amounts of transfected GFP plasmids correlates with increasing CRE-DOG activity, but beyond a certain point this correlation becomes negative (Figure 3.6E). This change is likely due to saturation of GBPs by excessive GFP level, resulting in inhibition of complex formation, as also observed in the transcription system (Figure 2.2E).

We explored the possibility of conferring CRE-DOG^{OPT} with drug inducibility, which would be useful for temporal control of GFP-dependent recombination. We fused the DHFR destabilizing domain (DD) (Iwamoto et al., 2010) to the N- or C-terminus of CRE-DOG^{OPT} components. These DD fusions were then assayed in reporter screens as pairwise combinations for those pairs that give recombination activity only in the presence of both GFP and TMP, a small molecule that stabilizes DHFR-DD (Iwamoto et al., 2010). The best performing pair was found to include the DHFR DD fusion to the C-terminal end of the N-CretrcintG fragment, and is named DD CRE-DOG^{OPT} (Figure 3.6A and F). DD CRE-DOG^{OPT} activity displays very low background activity without GFP and TMP, shows slight background with GFP addition, and is dependent on TMP dosage (Figure 3.6F). Thus, we isolated a tool that can potentially be applied for GFP- and drug inducible control of specific cell types.

Utility of CRE-DOG for *in vivo* studies

We tested whether CRE-DOG could be used to retrofit transgenic GFP reporter lines for cell-specific manipulations (Figure 3.7 and 3.8). We first tested CRE-DOG^{OPT} activity in the Tg(CRX-GFP) line. In mature retinas, Tg(CRX-GFP) expresses GFP strongly in all

Figure 3.7 Retrofitting transgenic GFP mouse lines for GFP-dependent Cre recombination of specific cell populations. (A) Schematic of electroporation experiment in (B-C). CAG-nlacZ encodes n-βgal is used as a marker for electroporation. CAG-N-CretrcintG and C-CreintG gives CRE DOG^{OPT}. Floxed reporter is CAG-loxP-Neo-loxP-DsRed (CALNL-DsRed) in (B) and ChAG-loxP-TagBFP-loxP-mCherry (ChAGLtBFPL-mCherry) in (C). (B) CRE-DOG^{OPT} activity in Tg(CRX-GFP) reporter line. ONL, outer nuclear layer. INL, inner nuclear layer. Scale bar, 30 μm. (C) CRE-DOG^{OPT} activity in Tg(PROX1-GFP) reporter line. Scale bar, 20 μm.

Figure 3.7 (Continued)

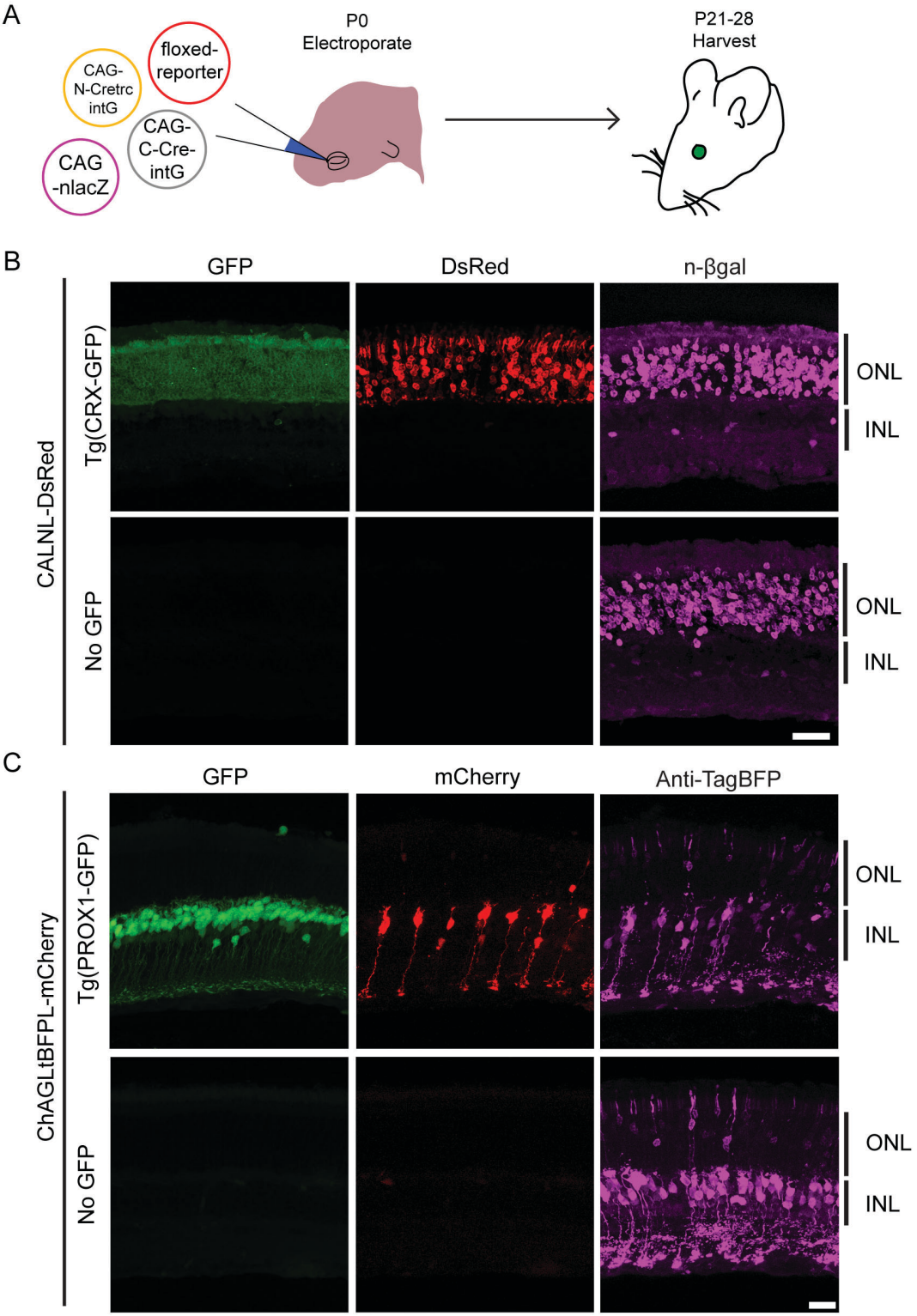
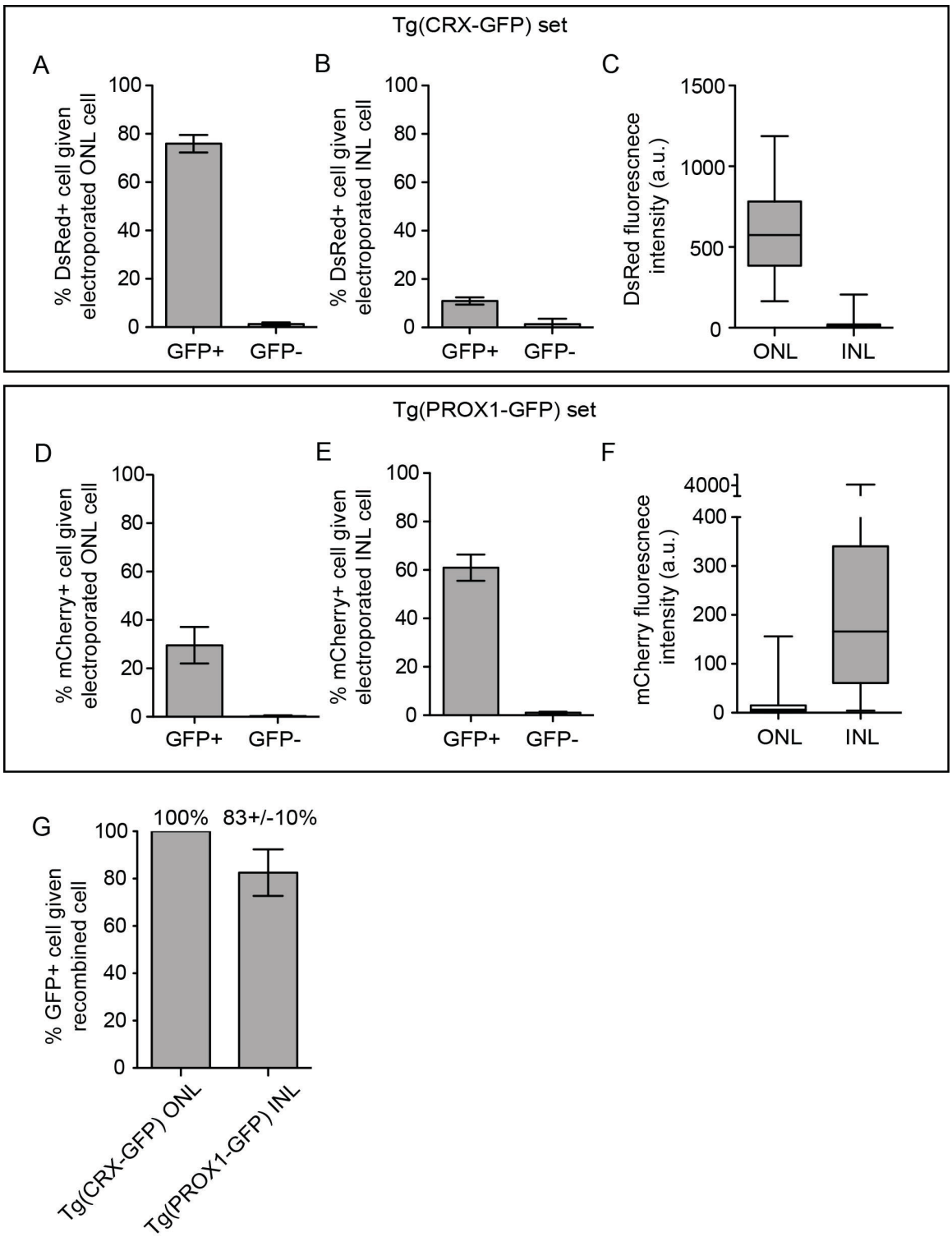


Figure 3.8. Quantification of electroporation experiment in Figure 3.7.

(A-B) Efficiency and GFP-dependency of CRE-DOG^{OPT} activity in electroporated Tg(CRX-GFP)+ or – retinas. Electroporated ONL (A) or INL (B) cells are defined by expression of n-βgal+ cells (from co-electroporated CAG-nlacZ plasmids). n=4 retinas per condition in (A). n=3 retinas per condition in (B). (C) Quantification of floxed DsRed expression in Tg(CRX-GFP)+, DsRed+ ONL and INL cells. n=28 ONL cells and 15 INL cells. (D-E) Efficiency and GFP-dependency of CRE-DOG^{OPT} activity in electroporated Tg(PROX1-GFP)+ or – retinas. Electroporated ONL (D) or INL (E) cells are defined by staining for Anti-TagBFP (TagBFP is from ChAG-loxP-TagBFP-loxP-mCherry cassette. Antibody cross-reacts with mCherry too). (F) Quantification of floxed mCherry expression in Tg(PROX1-GFP)+, mCherry+ ONL and INL cells. n=65 ONL cells and 67 INL cells. (G) GFP specificity of ONL DsRed (Tg(CRX-GFP) set) or INL mCherry (Tg(PROX1-GFP) set) cells. n=4 retinas per condition, at least 50 DsRed+ or mCherry+ cells counted per retina. All bar graphs show mean +/- standard deviation. All box plots show median and whiskers indicate minimum to maximum values

Figure 3.8 (Continued)



photoreceptors and weakly in INL cells, primarily bipolar cells (Samson et al., 2009). Electroporation of CRE-DOG^{OPT} along with a floxed DsRed reporter into postnatal Tg(CRX-GFP) retinas resulted in strong labeling of photoreceptors and occasional labeling of INL cells (Figure 3.7A and B). The observed DsRed pattern is GFP-dependent, as electroporated GFP- retinas show little to no DsRed labeling in both the ONL and INL (Figure 3.7B). Also, GFP immunostaining revealed that 100% of DsRed cells labeled in the ONL were GFP positive (Figure 3.8G). A considerable amount of retinal diversity lies in the bipolar cell class. We explored whether it is possible to target reporter lines expressing GFP in this cell class by electroporation. We used a hybrid Chx10-beta actin minimal promoter fusion element to drive strong expression of CRE-DOG^{OPT} components in bipolar cell types of the Tg(PROX1-GFP) line, and assayed Cre activity with a loxP-TagBFP-STOP-loxP-mCherry reporter. This reporter expresses TagBFP in the absence of Cre activity and switches to mCherry expression following Cre-mediated recombination. In contrast to Tg(CRX-GFP), we observed Tg(PROX1-GFP)-dependent expression of mCherry in bipolar cells, as well as sparse labeling of photoreceptors, and rare labeling of Muller glia and amacrine cells (Figure 3.7C). Although only ~83% of mCherry+ cells were actually GFP+ (Figure 3.8G), analysis of electroporated GFP- retinas indicated that the system was absolutely GFP-dependent in this assay (Figure 3.8 D-F). Thus, we speculate that the labeled cell types reflect developmental expression of Tg(PROX1-GFP) in progenitor cells that gives rise to photoreceptors and/or Muller glia and amacrine cells, or transient GFP expression in newly born cell types in these categories. Altogether, these results establish the utility of CRE-DOG for labeling and manipulating specific cell types in transgenic GFP animals.

Delivery by AAV

The delivery of CRE-DOG^{OPT} to the nervous system may involve transient techniques such as electroporation or viral infection. Recombinant adeno-associated viruses (rAAV) have become popular viral vectors for delivering transgenes to the mammalian nervous system, due to their high infectivity of postmitotic neurons and low toxicity (Betley and Sternson, 2011). We tested for CRE-DOG^{OPT} activity with the components controlled by the ubiquitous human elongation factor 1-α (EF1α) promoter, and delivered in the form of rAAVs. Due to size limit of rAAV vectors we had to encode the two components of CRE-DOG^{OPT} in separate rAAV vectors. We found that co-injection of rAAV-CRE-DOG^{OPT} with rAAV-EF1α-GFP, and a Cre-dependent FLEX-tdTomato (tdT) reporter (Atasoy et al., 2008) resulted in tdT expression in the mouse retina (Figure 3.9, top row). In contrast, replacement of GFP with the distantly related ZsGreen fluorescent protein (Matz et al., 1999) gave little to no tdT expression (Figure 3.9, bottom row). Thus, we demonstrated that CRE-DOG^{OPT} could be delivered to the mouse nervous system via rAAV.

Discussion

Here, we demonstrated that a high signal to noise GFP-dependent Cre recombinase could be obtained by fusing pairs of GBPs to the N- and C-terminal portions of Cre. GFP-specificity and overall activity is promoted by including portions of the VMA intein regions at the junction of the GBP and Cre domains. The optimized pairs of GBP-Cre fusions that include the intein sequences now enable the use of GFP to directly induce recombination of loxP-based cassettes. The immediate beneficiaries of these constructs

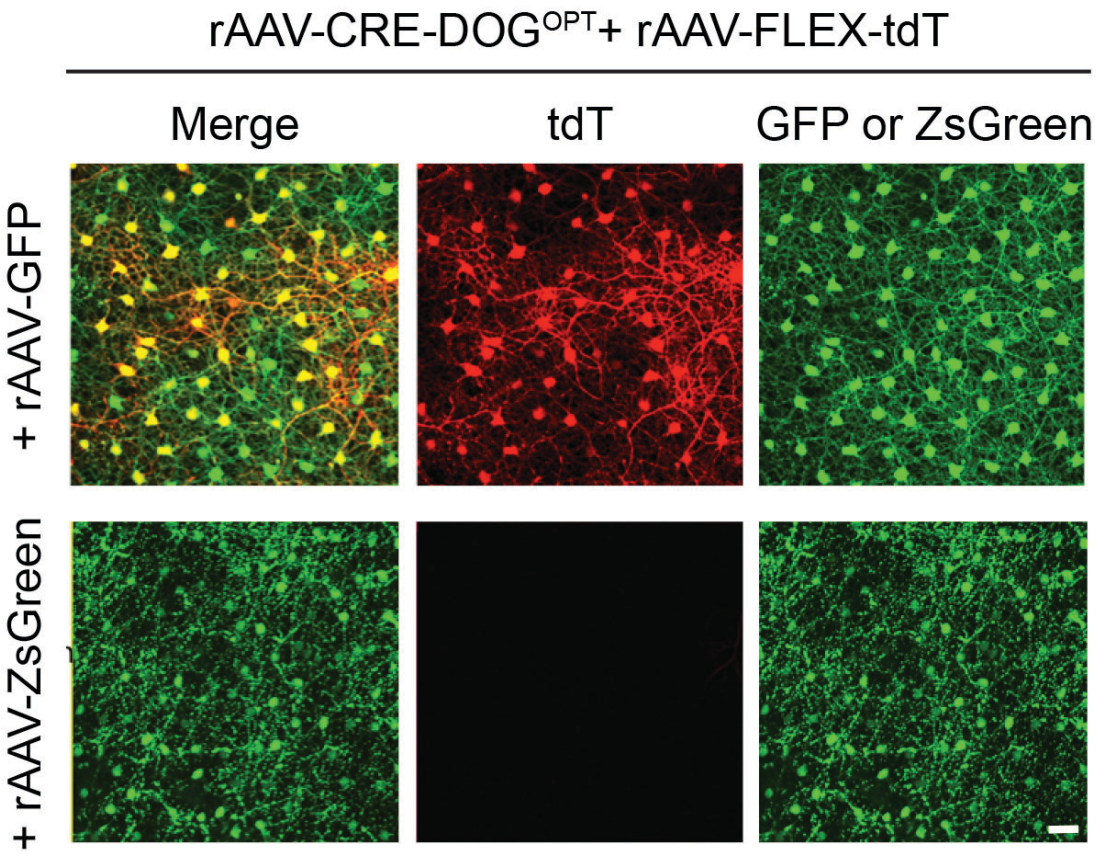


Figure 3.9. Delivery of CRE-DOG^{OPT} to the mouse nervous system by rAAV. Co-injection of rAAV (serotype 2/8) encoding EF1α-driven N-CretrcintG and C-CreintG and rAAV-FLEX-tdT into the P0 murine retina along with either rAAV encoding EF1α-driven GFP (top row) or ZsGreen (bottom row). Retinas were harvested between P21-30. Top row: GFP-dependent activation of tdT in the horizontal cell layer of the retina. Bottom row: little background CRE-DOG^{OPT} activity was observed. Note: ZsGreen was found to aggregate heavily in the retina. Images representative of at least 3 retinas per condition. Scale bar, 30 μm.

are those studying model organisms, with access or are in possession of transgenic GFP lines – these scientists will be able to retrofit existing transgenic GFP lines for cell-specific gene manipulation. CRE-DOGs may further be applied as components in a variety of plausible synthetic circuits. To our knowledge, this work is the first to demonstrate the utility of artificially-derived binding proteins (ex. nanobodies, intrabodies, etc) for constructing protein-inducible split enzyme systems.

The immediate delivery method of CRE-DOG^{OPT} to the mouse nervous system would be as DNA plasmids or viruses. In the longer term, we are building transgenic mouse lines expressing CRE-DOG^{OPT} components from a bicistronic IRES cassette for genetic crosses to transgenic GFP lines. The efficiency of CRE-DOG^{OPT} is currently limited by the fact that we have to encode each component in a separate rAAV. In the long term it would be beneficial to develop strategies to encode both components in a single rAAV vector, perhaps by linking the two components in a bicistronic cassette with 2a peptides, which are much shorter in DNA length than IRES elements.

Although split intein elements were found to be important for mediating the GFP-dependency of the system, protein splicing did not seem to play a major role, or any role at all. It is unclear why the inclusion of intein elements reduced background activity of CRE-DOG, but one could speculate that the extra elements influenced the orientation or conformation of the split protein fragments such that non-specific CRE-DOG assembly was inhibited in the absence of GFP. Whether split intein elements would also aid in

reducing background activity of other ligand-inducible split protein systems would be an interesting issue to address in the future.

Unlike the T-DDOGs described in Chapter 2, we did not detect any adverse side effects in cells overexpressing CRE-DOG. Further, the irreversibility of the system holds the advantage of enabling stable expression of transgene in cell types regardless of GFP level. However, there are also disadvantages of CRE-DOGs, and there are aspects of its use that must be taken into consideration. As with the introduction of any exogenous construct, the efficiency of CRE-DOG activity depends upon the effectiveness of gene delivery. Another concern with CRE-DOG is the irreversibility of the system. Although it was not found to be a problem in our experiments, the low but existent background for the current CRE-DOG is expected to lead to accumulation of background signal with longer incubation times. This can be overcome by incorporating temporal control over the system, such as using ERT2 or DHFR domains fused to CRE-DOG^{OPT}. Indeed, testing of the DD-CRE-DOG^{OPT} system is underway.

Another plausible issue with CRE-DOG, as is also the case with T-DDOG, will be sensitivity to GFP levels in excess of the split components. When GFP is in excess to CRE-DOG components, it will reduce recombination rate, due to titration of the GBP components. However, we note that many transgenic GFP lines available to the mouse at least are very weak in GFP expression, and GFP expression has to be revealed by antibody staining in many cases. Nonetheless, the Cre-dependent recombination was fairly efficient in such lines, indicating that these lines will be sufficient for many

experiments. We are currently testing rAAV delivery of CRE-DOG^{OPT} across the mouse nervous system. The data thus far indicate that CRE-DOG^{OPT} works across a number of transgenic GFP lines in the brain, retina and spinal cord, and that the system will work for a wide range of GFP levels (data not shown). Thus, we anticipate CRE-DOG^{OPT} to be broadly applicable for cell-specific manipulation of GFP labeled cells across transgenic mouse lines.

Materials and Methods

Animals – All animal experiments performed were approved by the Institutional Animal Care and Use Committee at Harvard University. Time pregnant CD1 (Charles River Breeding Laboratories, Boston, MA) were used for the electroporation experiment. Tg(CRX-GFP) (Samson et al., 2009) and Tg(PROX1-GFP) (gensat.org) were bred in C57/BL6J background. For experiments they are crossed to CD1, C57/BL6J or litter mates to generate pups.

Molecular biology and CRE-DOG screens – Using standard cloning techniques, coding sequences of GBPs were fused to those of split Cre and split intein fragments in many possible configurations, and the recombinants were inserted into a pCAG vector (Niwa et al., 1991). N- and C-terminal split Cre fragments used in CRE-DOG correspond to residue 19-59 and residue 60-343 of full-length Cre, respectively.

rAAV injections- rAAV (serotype 2/8) encoding EF1α-driven N-CretrcintG and C-CreintG components of CRE-DOG^{OPT} were co-injected with rAAV encoding EF1α-driven GFP or ZsGreen into the postnatal P0 mouse retina. Injected retinas were

harvested in adulthood, between P21-30. Injected rAAV titer were in the range of 10^{13}-10^{15} genome copies per ml.

In vitro **luciferase assays** –Plasmids encoding CAG-driven GFP, N- and C-terminal GBP-VMA-SplitCre fusion proteins were transfected via polyethyleneimine (PEI) into 293T cells along with plasmids encoding CALNL-luc2 and Renilla luciferase. Between 50-70 ng of total DNA were transfected into single wells of 48 well plates. Cells were harvested 24 hours later for Dual Luciferase Assay (Promega). All transfections, except for the dosage curve, were done at a 1:1:1 (GFP:N-CreintG:C-CreintG) plasmid molar ratio.

Western Blot: 293T cells were seeded onto 6 well plate and transfected with CAG-driven N-CreintG and C-CreintG-FLAG, along with CAG-GFP or empty CAG vector. A CALNL-DsRed reporter was included in some cases to confirm GFP-dependent transfection. 24 hours post-transfection, transfected 293T cells were lysed in RIPA buffer (150mM NaCl, 1.0% NP40, 0.5% Na Deoxycholate, 0.1% SDS, 50mM Tris, pH 8.0) for lysate supernatant load or 4x Laemmeli buffer plus beta-mercaptoethanol. Samples were ran on SDS PAGE and then subject to Western blot. Anti-FLAG was used at 1:1000 dilution.

In vivo **electroporations –** Postnatal day 0 to 3 (P0-P3) mouse pups were electroporated *in vivo* as described previously (Tang et al., 2013). DNA solutions (1-1.5μg/μl) were injected through the sclera and into the subretinal space of the mouse retina. pCAGEN was used as an empty vector substitute for excluded plasmids. P0 CD1 retinas were electroporated with plasmids encoding CAG-driven CRE-DOG components and nuclear

β-galactosidase (n-βgal, an electroporation marker expressed from CAG-nlacZ plasmid) along with CALNL-DsRed and different promoter-GFP constructs. Electroporated retinas were harvested at P14, immunostained for n-βgal and then imaged on a Zeiss LSM780 confocal microscope. At P2-3, Tg(PROX1-GFP) retinas were electroporated with plasmids encoding CRE-DOG components and ChAG-loxP-TagBFP-loxP-mCherry; retinas were harvested between 3-4 weeks of age for CALNL-DsRed. Wherever applicable, retinas were immunostained with anti-GFP or anti-DsRed to visualize processes.

Retinal Histology: Isolated mouse retinas were fixed at room temperature for 30 minutes in 4% paraformaldehyde/PBS solution. Retinas were then transferred to 30% sucrose in PBS, and subsequently into a 50/50 mixture of 30% sucrose/PBS and OCT for sectioning. 20 μm retinal cryosections were cut on a Leica CM3050 cryostat (Leica Microsystems), using disposable blades.

Retinal immunohistochemistry: Retinal cryosections were incubated in blocking solution (5% normal donkey serum, 0.1% Triton-X in PBS) for 1 hour and stained for primary antibody overnight at 4 degrees Celsius. Immunostained cryosections were washed three times in 0.1% Triton-X in PBS and stained for secondary antibodies in blocking solution for 2 hours at room temperature. Slides were then washed in 0.1% Triton-X in PBS and mounted for imaging in Fluoromount-G (Southern Biotechnology Associates; 0100-01).

Statistical Analysis - Two-tailed Student's t test assuming unequal variance was used for all comparisons. $p > 0.05$ is judged as statistically significant.

References

Anraku, Y., Mizutani, R., and Satow, Y. (2005). Protein splicing: its discovery and structural insight into novel chemical mechanisms. IUBMB Life *57*, 563-574.

Atasoy, D., Aponte, Y., Su, H.H., and Sternson, S.M. (2008). A FLEX switch targets Channelrhodopsin-2 to multiple cell types for imaging and long-range circuit mapping. J Neurosci *28*, 7025-7030.

Betley, J.N., and Sternson, S.M. (2011). Adeno-associated viral vectors for mapping, monitoring, and manipulating neural circuits. Hum Gene Ther *22*, 669-677.

Chalfie, M. (2009). GFP: Lighting up life. Proc Natl Acad Sci U S A *106*, 10073-10080.

Chalfie, M., Tu, Y., Euskirchen, G., Ward, W.W., and Prasher, D.C. (1994). Green fluorescent protein as a marker for gene expression. Science *263*, 802-805.

Gill, G., and Ptashne, M. (1988). Negative effect of the transcriptional activator GAL4. Nature *334*, 721-724.

Gong, S., Zheng, C., Doughty, M.L., Losos, K., Didkovsky, N., Schambra, U.B., Nowak, N.J., Joyner, A., Leblanc, G., Hatten, M.E*., et al.* (2003). A gene expression atlas of the central nervous system based on bacterial artificial chromosomes. Nature *425*, 917-925.

Heintz, N. (2004). Gene expression nervous system atlas (GENSAT). Nat Neurosci *7*, 483.

Iwamoto, M., Bjorklund, T., Lundberg, C., Kirik, D., and Wandless, T.J. (2010). A general chemical method to regulate protein stability in the mammalian central nervous system. Chem Biol *17*, 981-988.

Jullien, N., Sampieri, F., Enjalbert, A., and Herman, J.P. (2003). Regulation of Cre recombinase by ligand-induced complementation of inactive fragments. Nucleic Acids Res *31*, e131.

Matz, M.V., Fradkov, A.F., Labas, Y.A., Savitsky, A.P., Zaraisky, A.G., Markelov, M.L., and Lukyanov, S.A. (1999). Fluorescent proteins from nonbioluminescent Anthozoa species. Nat Biotechnol *17*, 969-973.

Nakai, J., Ohkura, M., and Imoto, K. (2001). A high signal-to-noise Ca(2+) probe composed of a single green fluorescent protein. Nat Biotechnol *19*, 137-141.

Niwa, H., Yamamura, K., and Miyazaki, J. (1991). Efficient selection for high-expression transfectants with a novel eukaryotic vector. Gene *108*, 193-199.

Samson, M., Emerson, M.M., and Cepko, C.L. (2009). Robust marking of photoreceptor cells and pinealocytes with several reporters under control of the Crx gene. Dev Dyn *238*, 3218-3225.

Shaner, N.C., Steinbach, P.A., and Tsien, R.Y. (2005). A guide to choosing fluorescent proteins. Nat Methods *2*, 905-909.

Shimomura, O., Johnson, F.H., and Saiga, Y. (1962). Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, Aequorea. J Cell Comp Physiol *59*, 223-239.

Siegert, S., Scherf, B.G., Del Punta, K., Didkovsky, N., Heintz, N., and Roska, B. (2009). Genetic address book for retinal cell types. Nat Neurosci *12*, 1197-1204.

Tang, J.C., Szikra, T., Kozorovitskiy, Y., Teixiera, M., Sabatini, B.L., Roska, B., and Cepko, C.L. (2013). A nanobody-based system using fluorescent proteins as scaffolds for cell-specific gene manipulation. Cell *154*, 928-939.

Tsien, R.Y. (1998). The green fluorescent protein. Annu Rev Biochem *67*, 509-544.

Tyszkiewicz, A.B., and Muir, T.W. (2008). Activation of protein splicing with light in yeast. Nat Methods *5*, 303-305.

Vila-Perello, M., and Muir, T.W. (2010). Biological applications of protein splicing. Cell *143*, 191-200.

Yizhar, O., Fenno, L.E., Davidson, T.J., Mogri, M., and Deisseroth, K. (2011). Optogenetics in neural systems. Neuron *71*, 9-34.

Chapter Four

Detection and Manipulation of Antigen-Expressing Cells with Destabilized Binding Proteins

Detection and Manipulation of Antigen-Expressing Cells with Destabilized Binding Proteins

Authors: Chung Yiu Jonathan Tang[1,2], Eugene Drokhlyansky[1,2], Constance L. Cepko[1,2]

[1]Howard Hughes Medical Institute

[2]Departments of Genetics and Ophthalmology

Harvard Medical School, Boston, MA 02115, USA

This chapter contains the manuscript titled "Detection and Manipulation of Antigen-Expressing Cells with Destabilized Binding Proteins", in preparation. It is modified to better fit the style of this dissertation. Jonathan C.Y. Tang is co-first author on this manuscript. J.C.Y.T initiated and coordinated all aspects of the project. J.C.Y.T and E.D. conducted all experiments and analysis, C.L.C. supervised the project. J.C.Y.T wrote the paper with contributions from C.L.C. and E.D.

Summary

We developed a general strategy for using an input intracellular protein to control stability of an output protein of interest. We used the Green Fluorescent Protein (GFP) to show that antigen binding to a destabilized nanobody can induce protein stabilization and thus the activity of a variety of nanobody fusion proteins in mammalian cells. GFP can also control the spatial expression pattern of GFP nanobody fusion proteins in the mouse retina. The destabilizing mutations in the GFP nanobody scatter along the highly conserved framework of nanobodies. Transfer of these mutations to a nanobody targeting the HIV C-terminal domain (CTD) resulted in CTD-inducible stabilization of nanobody fusion protein. Thus, this work establishes a binary system using GFP to directly control output protein activity, and provides a platform for generating binding proteins that are stabilized by their antigens. Further, this work illustrates the potential for rapid design of intracellular protein-responsive sensors and effectors with a residue code that can be grafted across conserved binding protein scaffolds regardless of antigen identity.

Introduction

The ability to target specific cell populations based upon expression of an intracellular biomolecule, or on specific molecular modifications, would greatly facilitate studies of basic biology, as well as therapeutic applications. For example, in developmental biology and neuroscience there are many studies that focus on tracing and manipulating the activity of specific cell types defined by their unique molecular profiles. In biomedicine, such approaches could provide ways to selectively target and manipulate

healthy or diseased cells of interest. Genetic manipulation is routinely performed on model organisms for which transgenesis and transient gene delivery methods are well established. However, to expand the workspace of biologists it is essential that new genetic tools be developed so one can more rapidly and conveniently study any species of choice in great depth.

Binary expression strategies provide a powerful means of manipulating specific cell populations. In such strategies, a "driver" molecule is expressed under a cell-type specific promoter, and it interacts with a responder element to drive target gene expression. In one manifestation, the driver molecule could be a transcription factor that binds its cognate upstream activating sequence (UAS), resulting in transcription of a target gene under UAS control. Examples of this are the GAL4/UAS system (Brand and Perrimon, 1993), LexAop (Butala et al., 2009) and TetON/OFF systems (Schonig et al., 2010). The driver molecule could also be a site-specific DNA recombinase that recognizes its cognate binding sequences to induce DNA recombination events, leading to outcomes such as gene activation or deletion. Site-specific DNA recombinases have become the predominant driver molecules of choice in the mouse research community, with the most popular versions being the Cre/LoxP (Orban et al., 1992) and Flp/FRT systems (Dymecki, 1996).

Binary systems are powerful due to a number of reasons. First, since expression of the driver molecule is separated from that of the target gene, different cell-specific driver constructs can be combined with different responder-target gene constructs to perform a

wide variety of experiments. This modularity greatly reduces the number of genetic constructs needed to be generated for each experiment. Second, in the context of transgenic animals, the use of a default and innocuous driver molecule to create cell-specific driver lines makes it more likely that the resultant transgenic animals would be viable, with normal development and behaviors. Third, driver molecules can amplify the expression level of the target gene, as one driver molecule may catalyze or induce production of multiple target molecules.

Driver molecules are traditionally selected for their desired, natural biological activities, and are usually exogenous to the system of study. In Chapter 2, I propose that driver molecules can potentially be any intracellular product, exogenous or endogenous, as long as strategies exist to exploit their presence for "driving" desired molecular output. I demonstrated this using GFP as a proof-of-concept molecule. However, the dimerizer-based GFP-dependent systems described thus far were relatively tedious to apply, due in part because relatively large number of components are needed to drive target gene output. For example, in order for GFP to induce transcription of Cre, at least 3 components must be delivered and expressed in GFP-labeled cells in the T-DDOG system (Figure 2.5E-H), and at least 2 components have to be delivered in the CRE-DOG system to create Cre activity (Figure 3.7). Also, dimerizer systems suffered from the caveat that at concentrations in which the intracellular product of interest is in far excess of the split components, the activity of the system could be inhibited (Figure 2.2E; Figure 3.6E). An even simpler GFP-dependent system would be binary, whereby GFP interacts with a single component to directly induce activity of a protein of interest. Previously,

protein domains have been destabilized such that their stability and activity are dependent on binding to a small molecule or light, allowing the researcher to impose temporal control over activity of the destabilizable protein. (Banaszynski et al., 2006; Bonger et al., 2014). Here, I propose to use intracellular molecules as stabilizing ligands to enable spatial control of output protein activity with cellular precision.

Here, I describe a binary system using GFP to induce stabilization of a mutated GFP-binding nanobody (GBP1). This mutant nanobody could destabilize multiple fusion proteins tested here, and could be applied in the mouse for GFP-dependent and cell-specific regulation of protein expression. We generalized this approach by showing that the destabilizing mutations found in the GFP nanobody could be grafted onto another nanobody bound by the HIV C-terminal domain (CTD) protein. The approach taken here may be generalized to other artificially derived binding proteins for design of protein-responsive sensors and effectors.

Results

We propose to co-opt intracellular proteins as input signals to control the stability of destabilized binders such that the activity of an output protein, fused to the destabilized binder, will be directly controlled by interactions between the input protein and destabilized binder (Figure 4.1). Single chain antigen recognition domains derived from Camelid antibodies, or nanobodies, constitute an attractive class of protein binders for testing this idea, given their conserved protein backbone structure and potential to be selected to bind a variety of antigens in living cells (Muyldermans, 2013). To test

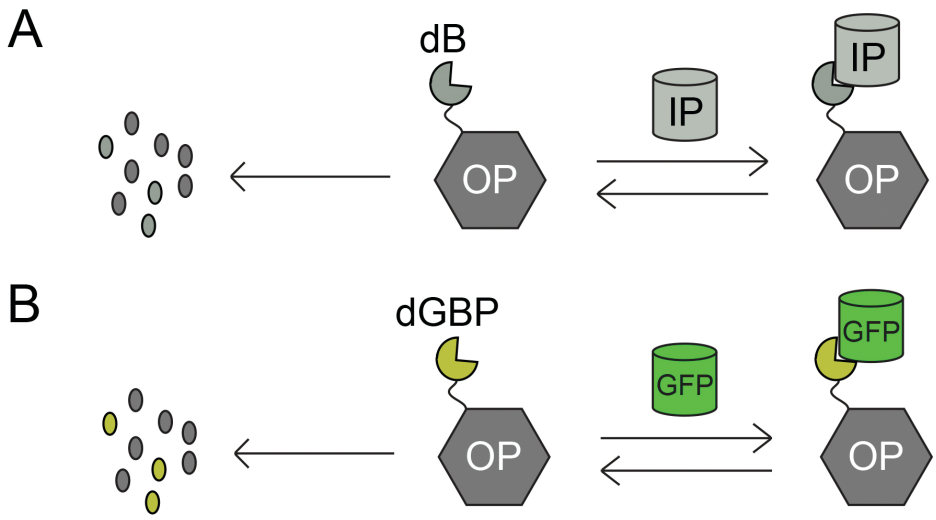
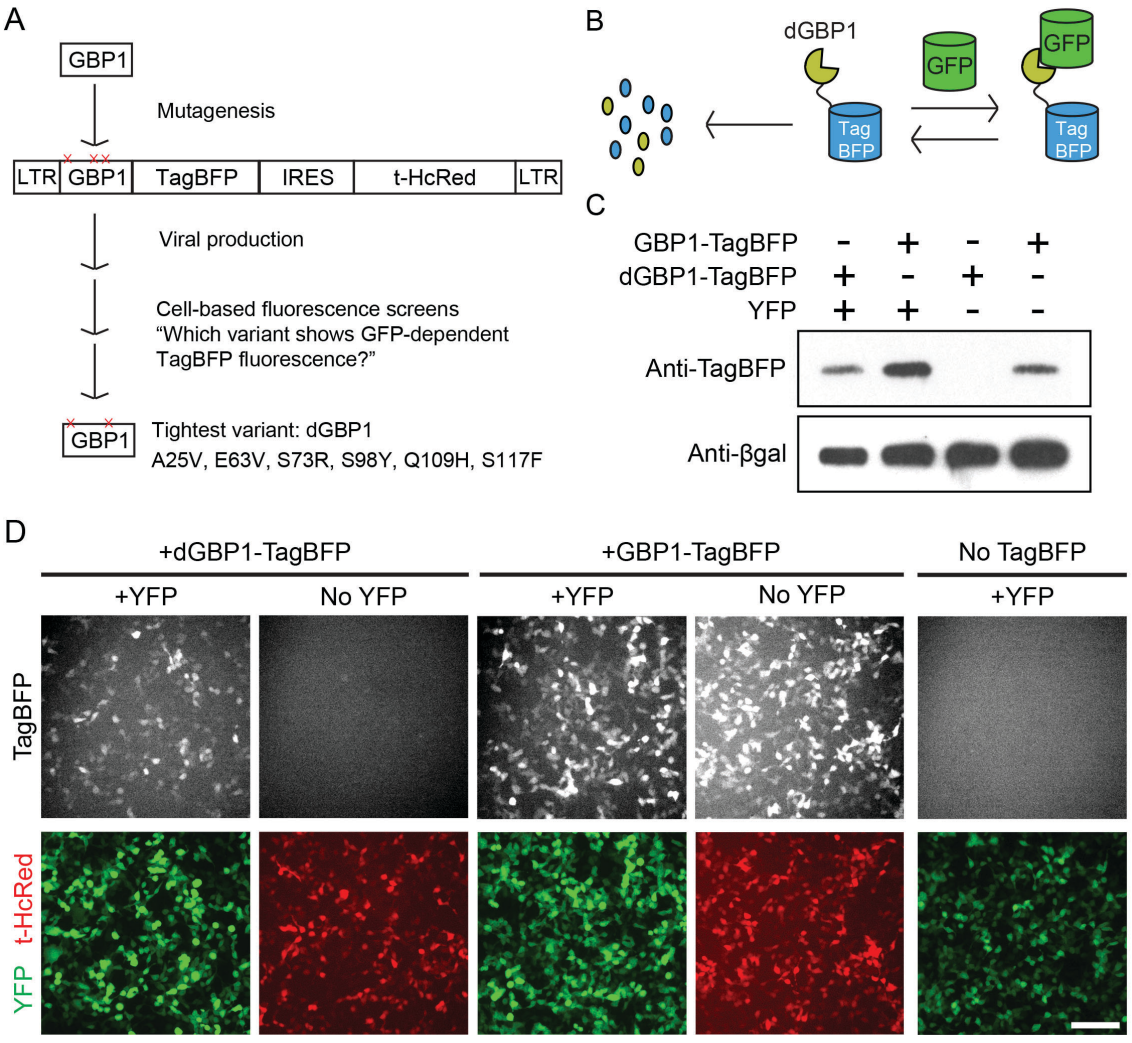


Figure 4.1. Concept of protein-inducible protein stabilization. (A) Output protein (OP) fused to a destabilized binder (dB) is degraded or inactive until binding of dB to input protein (IP) to form a complex. (B) GFP binding to a destabilized GFP binding protein (dGBP) induces stabilization and accumulation of the OP.

whether it is possible to modify nanobodies such that they are stabilized by expression of the target protein, we used the GFP-GBP1 complex (Kirchhofer et al., 2010) for proof-of-concept studies (Figure 4.1B and 4.2A). We mutagenized the GBP1 coding sequence and cloned the variants into a mouse leukemia virus (MLV) retroviral vector. GBP1 variants were inserted upstream and in frame to TagBFP, a blue fluorescent protein derived from *Entacmaea quadricolor* that bears little amino acid similarity to Aequorea-derived GFP and its derivatives (Subach et al., 2008). The mutant GBP1-TagBFP sequence was also part of a bicistronic cassette, in which the red fluorescent protein t-HcRed (Gurskaya et al., 2001) was expressed via an internal ribosomal entry site (IRES). 293T cells were infected with MLV viral particles encoding the GBP1-TagBFP variants, and the cells were subjected to fluorescence-activating cell sorting (FACS) for low TagBFP, HcRed+ population. Sorted cells were then infected with adeno-associated virus (AAV) expressing GFP, followed by FACS for cells with high TagBFP, HcRed+ fluorescence. Sorted cells were subjected to genomic extraction and PCR to isolate GBP1 variants. GBP1 variants were then screened individually for enhanced TagBFP expression in the presence of YFP, a GFP derivative. We were able to isolate a number of variants that showed YFP-dependent fluorescence. In contrast, wildtype GBP1-TagBFP showed strong blue fluorescence signals regardless of the presence of YFP (Figure 4.2C). Strikingly, many such variants showed aggregation of TagBFP in punctate regions within the cell in the absence of YFP, but became soluble in the cytoplasm when co-expressed with YFP (data not shown). This suggests that YFP binding prevents destabilization of GBP1-TagBFP variants, which aggregate when unstable.

Figure 4.2. Isolation of a highly destabilized GBP1 variant that can be stabilized in expression by GFP. (A) General scheme for isolating destabilized GBP1 variants. Mutagenized GBP1 were cloned into a Mouse Moloney Leukemia Viral (MMLV) vector, for which the expression cassette is flanked by Long Terminal Repeat (LTR) sequences for packaging into viruses. Cloned GBP1 variants are fused in frame to TagBFP. t-HcRed is linked to GBP1-TagBFP transcription and serves as an expression control red fluorescent protein. Neither TagBFP nor t-HcRed is expected to interact with GBP1. The tightest variant is called dGBP1. (B) Working model for GFP-regulated protein stabilization. dGBP1-TagBFP is degraded unless bound by GFP. (C-D) dGBP1, but not wildtype GBP1, destabilizes the TagBFP fusion partner in cell culture. 293T cells were transfected with plasmids encoding the GFP derivative YFP or filler plasmids along with the dGBP1-TagBFP or GBP1-TagBFP MMLV constructs. (C) Western blot for TagBFP and B-gal (a transfection control) shows GFP-dependent stabilization of both wildtype and destabilized GBP1. Harvested 2 days post-transfection. Representation of 3 independent protein gel loads. (D) Representative images showing GFP-dependent fluorescence of dGBP1-TagBFP. Cells were imaged 32 hours after transfection. Shown are representative images. t-HcRed (red) labels cells with the TagBFP fusion constructs. Scale bar, 100 μm. Consistent results were observed across triplicates and confirmed in at least four independent experiments.

Figure 4.2. (Continued)



Characterization of GBP1 mutants stabilized by GFP

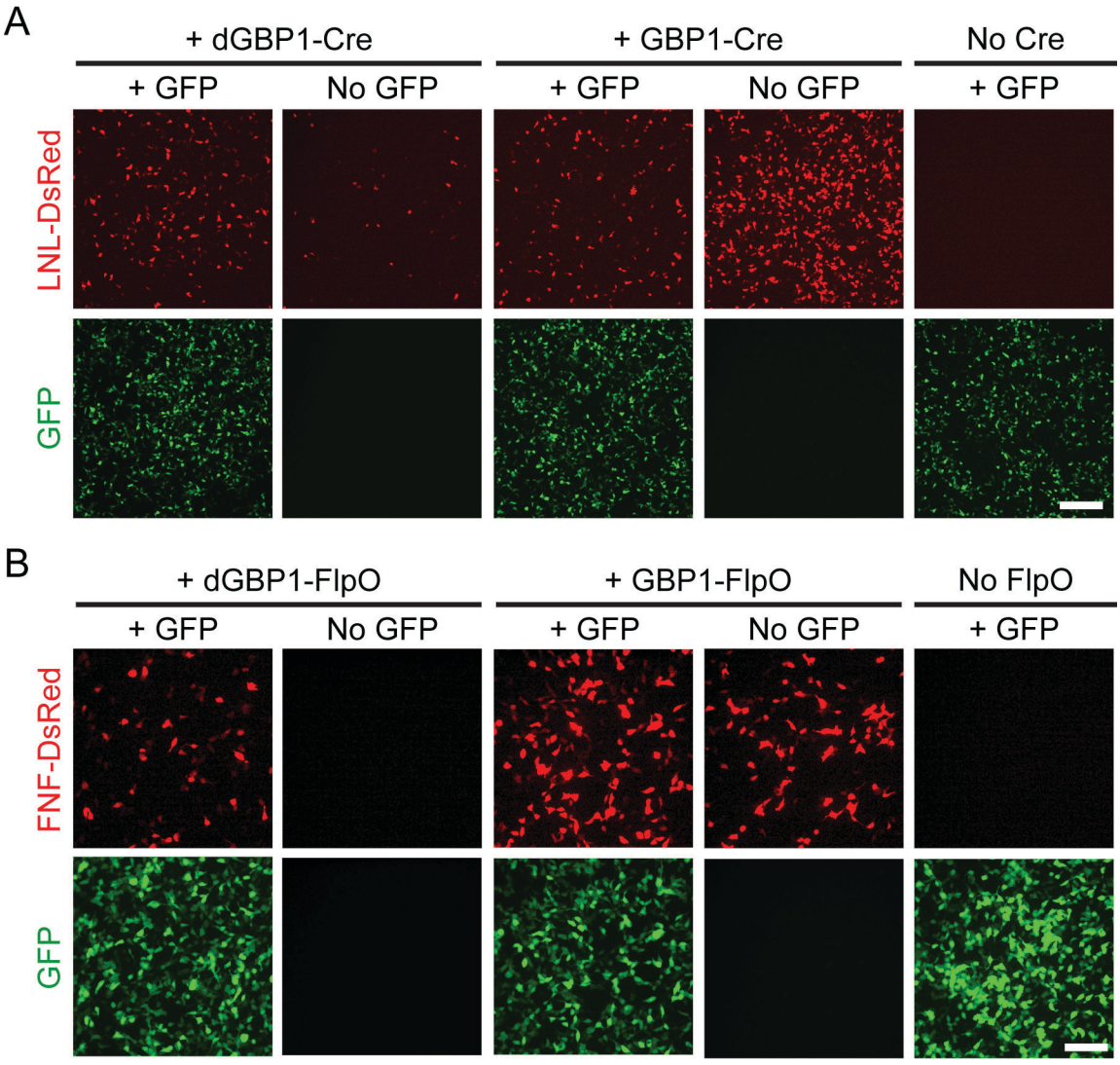
Amongst the many GBP1 variants that showed GFP-dependent stabilization, a variant carrying 6 amino acid changes (A25V, E63V, S73R, S98Y, Q109H, S117F) gave the lowest TagBFP fluorescence in the absence of GFP. We focused our efforts on this variant, which will hereafter be referred to as destabilized GBP1 (dGBP1). By fluorescence imaging and western blot analysis, we confirmed that dGBP1-TagBFP expression was induced by YFP (Figure 4.2C-D). Interestingly, we detected a possible enrichment of wildtype GBP1-TagBFP in the presence of YFP, suggesting that YFP could stabilize the expression of wildtype GBP1. Considering that we previously observed that GBP1-bearing T-DDOG could stabilize GFP expression *in vivo* (Figure 2.4E, Appendix I), it is possible that GBP1 and GFP mutually stabilizes each other.

Destabilized GBP1 can be used to control expression of multiple fusion protein partners in a GFP-dependent manner

The destabilizing effect of dGBP1 could be extended to fusion protein partners beyond TagBFP. We first tested for dGBP1 effects on site-specific recombinases Cre and Flp, which are commonly used as cell-specific driver molecules in mouse genetics (Orban et al., 1992) (Figure 4.3 and 4.4). We fused dGBP1 to Cre and codon-optimized Flp (Flpo) (Raymond and Soriano, 2007), creating dGBP1-Cre and dGBP1-Flpo, respectively. We tested for recombinase activity by the use of a DsRed reporter that was transcriptionally inactive until Cre or Flp-dependent excision of a transcription stop cassette flanked by loxP or FRT, respectively. Indeed, dGBP1-Cre and –Flpo had GFP-dependent recombinase activity in 293T cells (Figure 4.3). As the background recombination does

Figure 4.3. dGBP1 can confer GFP-dependency to multiple fusion protein partners. Plasmids encoding wildtype or destabilized GBP1 fusion to Cre (A) or Flpo (B) were transfected along with loxP-Neo-loxP-DsRed (A) or FRT-Neo-FRT-DsRed (B) reporter plasmids into 293T cells and analyzed for recombination to activate DsRed. (A) Scale bar, 200 μm. (B) Scale bar, 100 μm. Results were representation of triplicates per experiments in 3 independent experience.

Figure 4.3. (Continued)

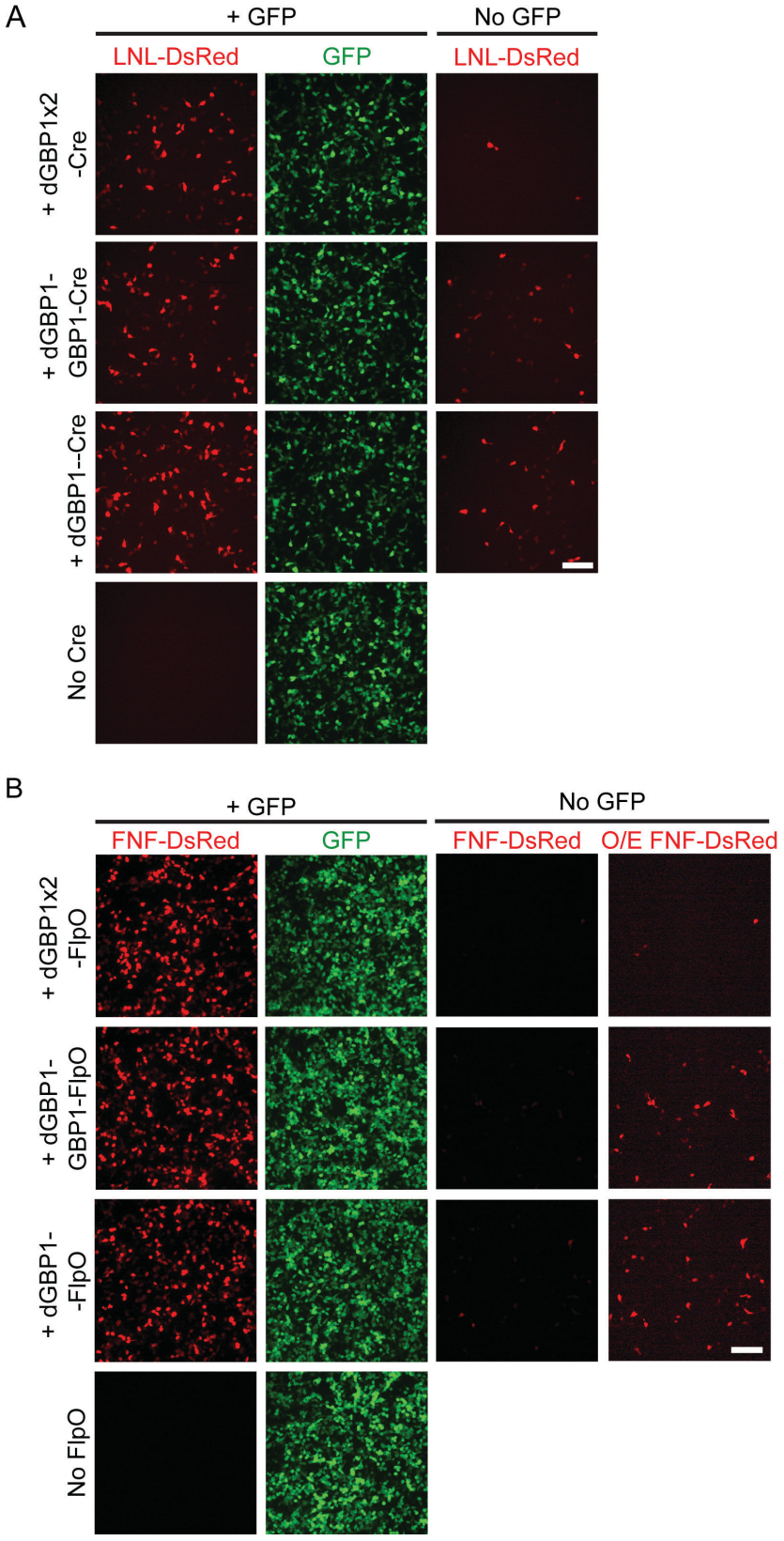


build up over time, especially with dGBP1-Cre, we explored strategies to reduce the background. We found that increasing the number of dGBP1 domains fused to Cre or Flpo was able to reduce GFP-independent recombination without noticeably affecting GFP-dependent recombination (Figure 4.4). In contrast, replacement of dGBP1 with wildtype GBP1 as one of the repeating unit gave relatively higher background signal (Figure 4.4). The best-performing construct over the assayed period was dGBP1x2-Flpo, with little to no accumulated background signal while the dGBP1-GBP1-Flpo and dGBP1-Flpo fusion construct accumulated a considerable amount of background recombination events (Figure 4.4). Taken together, these results indicate that dGBP1 is capable of conferring GFP-dependent control over different fusion protein partners, and that it is possible to optimize for background leakage by dGBP1 multimerization.

We found that GBP1 was also able to regulate the expression of the Discosoma-derived mCherry (Figure 4.5 and Table 4.1). We noticed that dGBP1-mCherry fusions did not disappear without GFP, as was observed for TagBFP. Instead, dGBP1-mCherry appeared to form aggregates within the cell. We exploited this property to use dGBP1-mCherry as a sensitized reporter to map the key residues involved in GBP1 destabilization, by comparing the level of fluorescence and aggregation of the fusion proteins in cells. Each mutation was tested for its necessity in creating aggregates, and it was found that S73, S98 and S117 were strongly involved in the destabilization effect. This observation was supported by sufficiency experiments, whereby the same mutations by themselves were able to destabilize mCherry (Figure 4.5 and Table 4.1). Interestingly, GFP addition was able to rescue the destabilization phenotype of all mutants.

Figure 4.4. Effect of dGBP1 multimerization on background recombinase activity. (A) 293T cells transfected with CAG-GFP, CAG-loxP-Neo STOP-loxP-DsRed (LNL-DsRed) and dGBP1 fusion Cre construct. Double dGBP1-Cre fusion (dGBP1x2-Cre) gave lowest background recombination activity. Image taken 22 hour post transfection. Data representative of triplicates and three independent experiments. Scale bar, 100 μm. (B) 293T cells transfected with CAG-GFP, CAG-FRT-Neo STOP-FRT-DsRed (FNF-DsRed) and dGBP1 fusion Flpo construct. Double dGBP1-Flpo fusion (dGBP1x2-Flpo) gave lowest background recombination activity. O/E: overexposed. Image taken 51 hour post transfection. Data representative of triplicates and three independent experiments. Scale bar, 100 μm

Figure 4.4 (Continued)



b

Figure 4.5. Mapping of key destabilizing mutations in dGBP1. Representative images of transfected 293T cells with wildtype and destabilized GBP1-mCherry and their variants. (A) dGBP1-mCherry fusion compared to variants where one of the six dGBP1 mutations was converted back to wildtype residue. (B) GBP1-mCherry fusion variants where one of the six dGBP1 mutations was introduced. Results representative of duplicate transfections. Imaged 34 hours post-transfection. Scale bar, 50um.

Figure 4.5 (Continued)

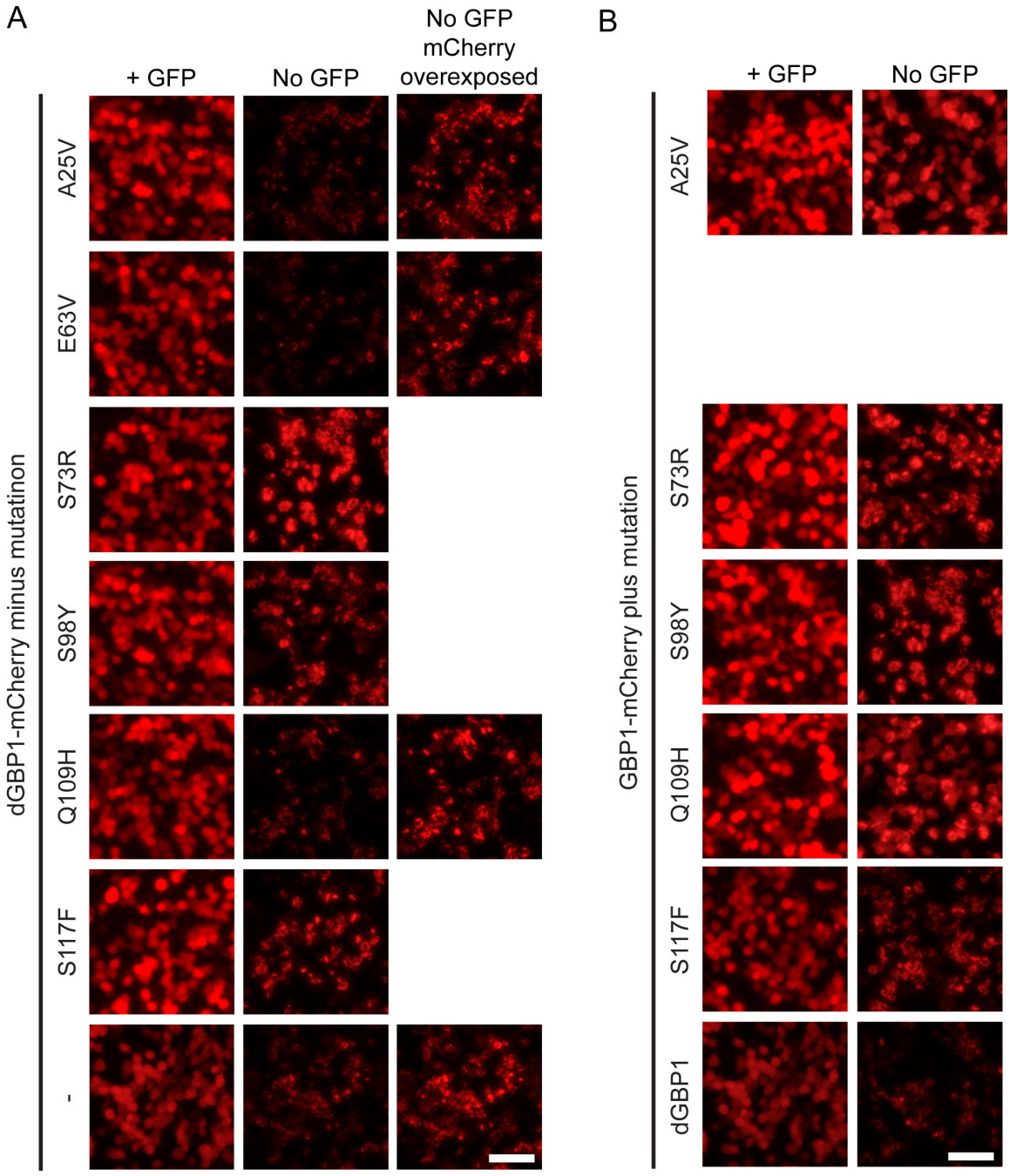


Table 4.1. Semi-quantitative summary of results from dGBP1 destabilizing residue mapping experiments

	No GFP		Plus GFP	
GBP1 plus mutations	**mCherry intensity**	**Aggregation**	**mCherry intensity**	**Aggregation**
A25V	+++	+	+++	-
E63V	N/A	N/A	N/A	N/A
S73R	++-	++	+++	-
S/C98Y	++-	+++	+++	-
Q109H	+++	++	+++	-
S117F	++	+++	++-	+
dGBP1	+	+++	++	+

	No GFP		Plus GFP	
dGBP1 minus mutations	**mCherry intensity**	**Aggregation**	**mCherry intensity**	**Aggregation**
A25V	-	+++	+++	+
E63V	-	+++	+++	-
S73R	+++	+++	+++	-
S/C98Y	++	+++	+++	+
Q109H	+	+++	+++	-
S117F	++	+++	++-	-
-	+	+++	++	-

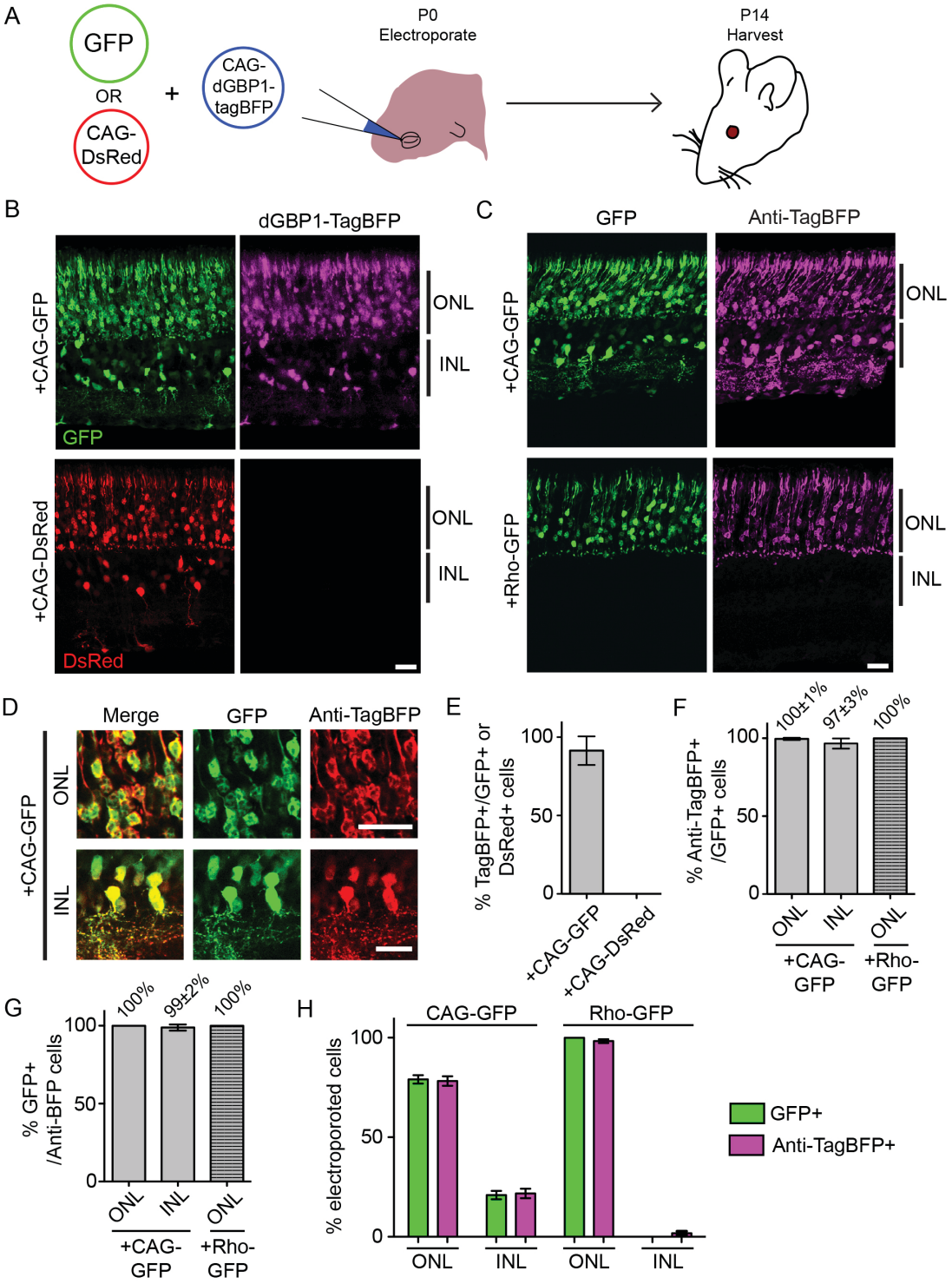
+ denote single score. The more score the more mCherry or more aggregation
- denote half a score.
N/A – not available.

GFP-dependent protein stabilization enables control of proteins in specific cell types in the mouse

We next tested whether the dGBP1 system could be applied *in vivo*. We used the mouse retina for such tests owning to the ease with which one can deliver plasmid DNA to the system and the availability of known cell-specific promoter elements for manipulating GFP expression pattern. We electroporated dGBP1-TagBFP into the postnatal mouse retina, along with different promoter-GFP constructs, and examined TagBFP expression upon tissue maturation (Figure 4.6). dGBP1-TagBFP was indeed destabilized in electroporated retinas without GFP expression, but was stabilized either broadly by CAG-GFP or exclusively in rod photoreceptors by Rho-GFP (Figure 4.6B, C, E and H). Replacement of GFP with DsRed resulted in loss of TagBFP fluorescence (Figure 4.6B and E). Strikingly, the efficiency of TagBFP activation and GFP co-localization of TagBFP positive cells were almost 100% (Figure 4.6F). In stark contrast, the efficiency of T-DDOG and CRE-DOG systems have never been observed to reach over 85% in similarly designed electroporation experiments (Chapter 2 and 3). This difference likely reflects that, with our new binary approach, fewer components are required for the desired GFP-dependent effect to occur. Further, the GFP specificity of TagBFP expressing cells is close to or at 100% (Figure 4.6G). Thus, these results showed that dGBP1 can be used as a tool to impose GFP-dependent control on fusion protein expression in the mouse, and may have greater efficiency in gene expression control over previous GFP-dependent systems.

Figure 4.6. GFP can control protein stabilization in a cell type-specific manner *in vivo.* (A) Schematic of electroporation experiment. Plasmid encoding promoter-GFP or CAG-driven DsRed is co-electroporated with CAG-driven dGBP1-TagBFP to the murine retina at postnatal day 0 (P0). Retinas were harvested at P14. (B) GFP (green), but not DsRed (red), induces accumulation of dGBP1-TagBFP (magenta) in the mouse retina. ONL, outer nuclear layer; INL, outer nuclear layer. (C) The expression pattern of immunostained TagBFP (magenta) can be altered by changing the GFP expression pattern with broadly-active (CAG) or rod photoreceptor-specific (Rho) promoters. Anti-TagBFP was not used in (A) as it cross-reacts with DsRed. (D) Co-localization of GFP (green) and Anti-TagBFP staining (red) from +CAG-GFP condition. (E-H) Quantification of electroporation results. (E) GFP-dependency of TagBFP expression. Counted cells from ONL. Plotted % TagBFP+ cells given GFP+ (from +CAG-GFP) or DsRed+ cells (from +CAG-DsRed). (F) Efficiency of GFP-dependent protein stabilization. Efficiency is % Anti-TagBFP+ cells given GFP+ cells. (G) GFP-specificity of system, as determined by % GFP+ cells given Anti-TagBFP+ cells. (H) dGBP1-TagBFP expression pattern closely matches that of GFP. All electroporated cells, as defined by GFP or TagBFP expression, are quantified across a 20 μm retinal section and represented as % of total number of cells counted. Scale bar is 20 μm. Graphs and values shown are as mean ± standard deviation. Sample size (retinas): n=3 for all conditions.

Figure 4.6. (Continued)



Generality of destabilizing mutations on other nanobodies

Since all six mutations found in dGBP1 are located within the non-hypervariable regions of the nanobody scaffold, and were not directly involved in antigen binding, we reasoned that it might be possible to simply transfer the mutations found in dGBP1 to other nanobodies for rapid generation of antigen-dependent protein sensors and effectors. In our initial tests, we found that although the six mutations could be used to destabilize all tested nanobody fusion proteins, we did not detect any clear antigen-dependent stabilization effects. In several cases, we did observe antigen-dependent changes in the fusion protein in terms of level and subcellular localization. However the changes were marginal (data not shown).

To understand why the six GBP1 destabilizing mutations could not be grafted for the desired effect, we studied the crystal structure of the nanobody-antigen complexes tested. We found that GFP bound to GBP1 via interactions with the CDR2 and CDR3 loops and with FR2 and FR3 (Figure 4.7B). This differs from the other nanobodies we initially tested, such as that of a GBP4-GFP complex (Figure 4.7C). In these other nanobodies, antigen binding occurs primarily at an elongated CDR3 loop region, pushing the antigen away from FR2, FR3 and CDR2. A survey through published structures revealed that some antigen-nanobody complexes do share a similar binding relationship as between GBP1 and GFP. Notably, the HIV C-terminal domain (CTD) nanobody (Nb) also has a relatively short CDR3 loop, leading to interactions with CTD at FR2, FR3 and CDR2, as did GBP1 with GFP (Figure 4.7A). Thus, we grafted the dGBP1 mutations to the CTD Nb (Figure 4.8). As expected, the 6 dGBP1 mutations destabilized CTD Nb-mCherry, or

Figure 4.7. Structural alignment of three antigen-nanobody complexes. (A-C) Two different views of crystal structure complexes of: (A) HIV C-terminal domain (CTD) (dark red) bound to its nanobody (CTD Nb) (light purple). (B) GFP (light green) bound to GBP1 (dark green). (C) GFP (orange) bound to GBP4 (wheat). Note the arrow denotes the end of the CDR3 loop in GBP1 in (B). GBP4 has an elongated loop, pushing the GFP antigen away from the backbone structure, whereas CTD Nb and GBP1 do not. (D) Superimposition of CTD/CTD Nb complex (dark red/light purple) on top of GFP/GBP1 complex (light green/dark green). (E) Superimposition of CTD/CTD Nb complex (dark red/light purple) on top of GFP/GBP4 complex (orange/wheat). (F) Superimposition of GFP/GBP1 complex (light green/dark green) on top of GFP/GBP4 complex (orange/wheat).

Figure 4.7 (Continued)

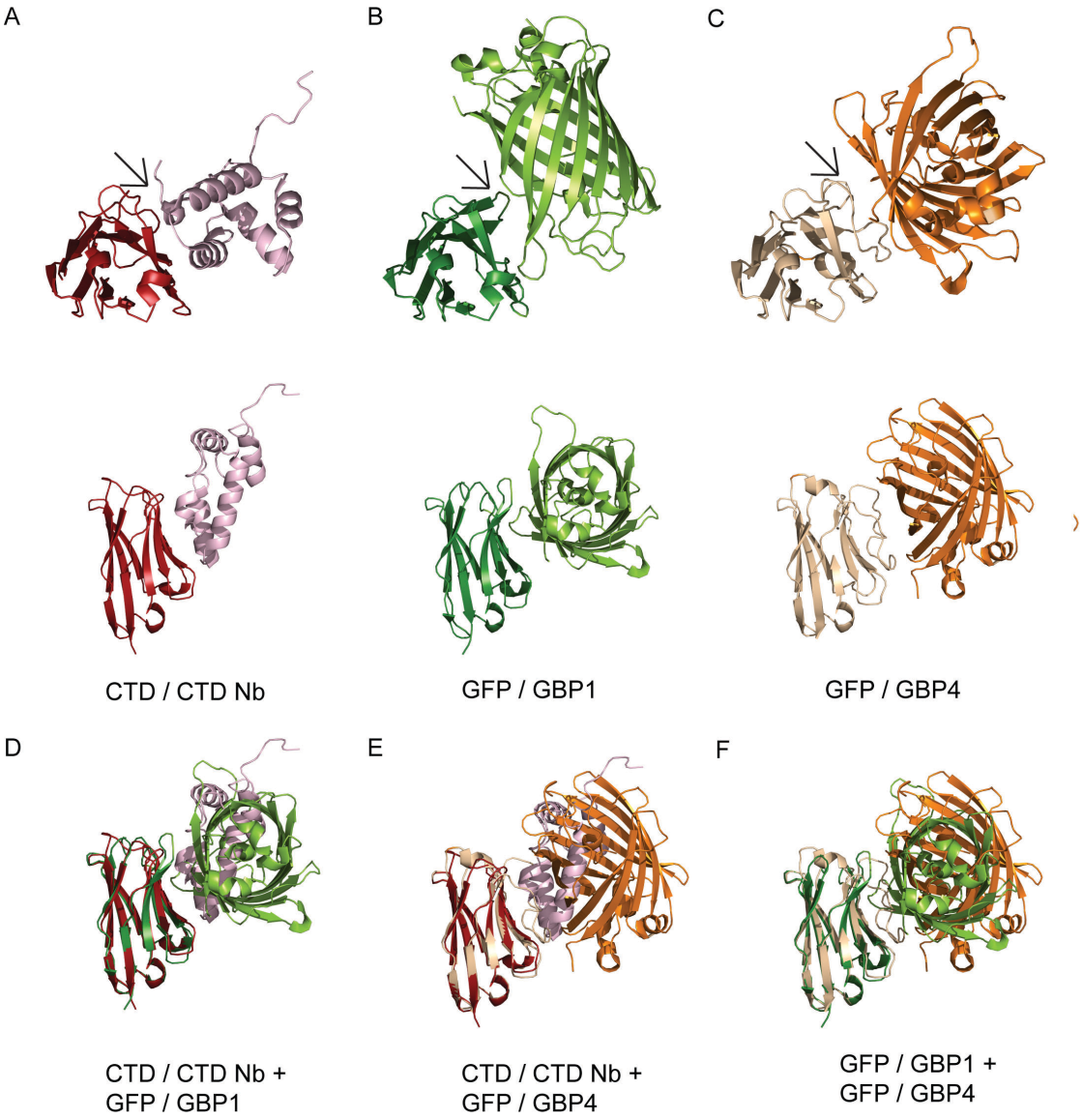
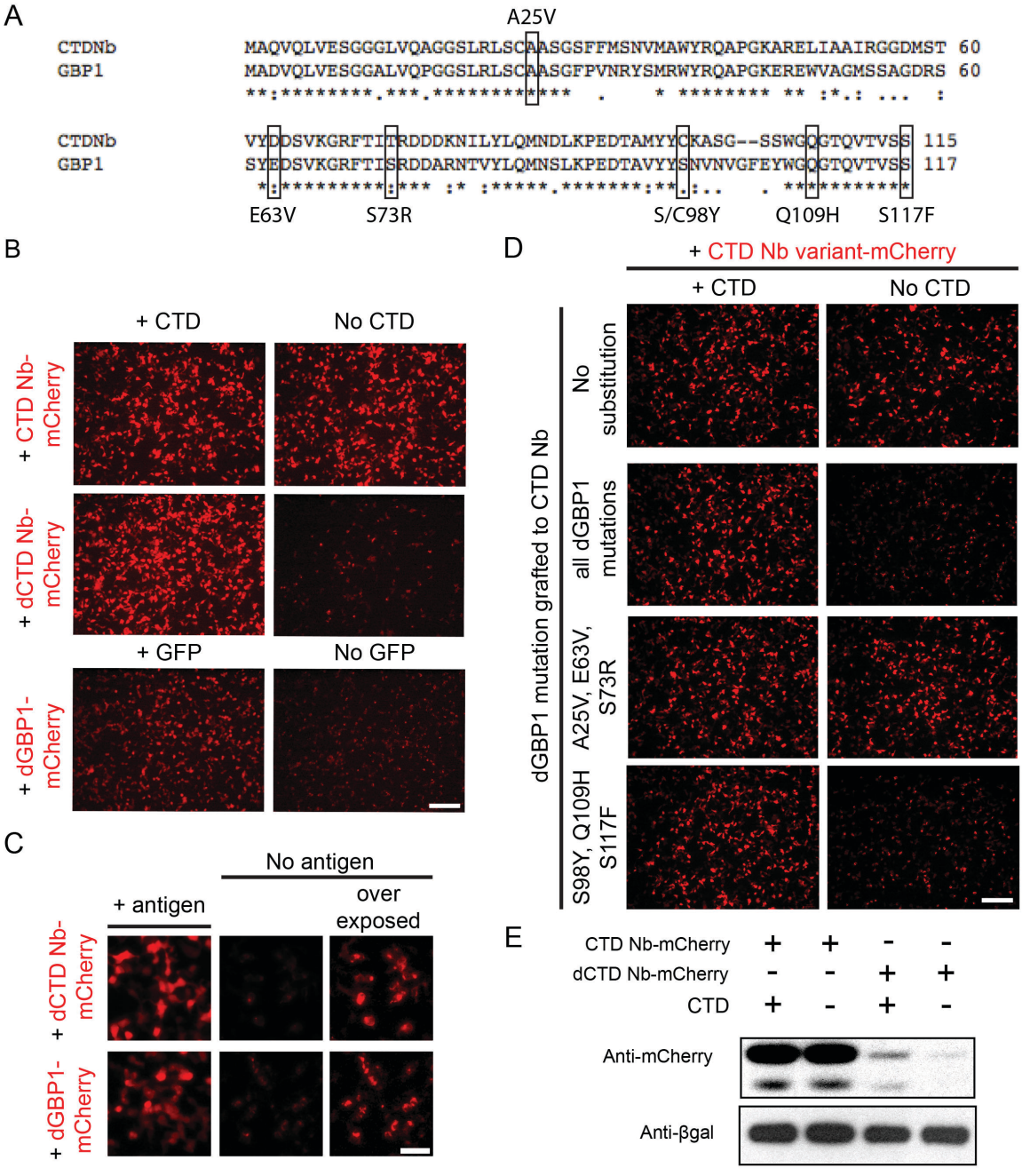


Figure 4.8. Destabilized CTD nanobody generated by grafting the destabilizing mutations from dGBP1. (A) Protein alignment of CTD nanobody (Nb) against GBP1. Black box highlights GBP1 residues mutated in dGBP1. Noted mutation positions correspond to dGBP1. Note that position 98 is originally cysteine in GBP1 as well, but was changed to serine in the version we used. (B) Transfer of dGBP1 mutations to CTD Nb generates destabilized CTD nanobody (dCTD Nb). Both dCTD Nb-mCherry and dGBP1-mCherry construct increases in fluorescence in the presence of its respective antigen. Images taken from transfected 293T cells,16 hours post-transfection. 1.5:1 DNA weight ratio of CAG-driven antigen plasmid to nanobody-mCherry plasmid. Scale bar, 100 μm. Consistent results were obtained in 3 independent experiments. (C) Close up view of cells from experiment in (B), taken 22 hour post transfection. Note aggregation of mCherry in "No Antigen" condition. For top panels, the antigen was CTD. For bottom panels, the antigen was GFP. Scale bar, 40 μm. (D) Sub-mapping of dGBP1 mutations involved in CTD Nb destabilizing phenotype. S98Y, Q109H and S117F appear to be involved in CTD Nb destabilization. Images taken 16 hours post-transfection. 1:1 DNA weight ratio of antigen to nanobody-mCherry plasmid transfected. Scale bar, 100 μm. All experiments were done in duplicates or triplicates, and consistent results were obtained in at least 2 independent experiments. (E) Western blot confirming CTD-inducible stabilization of dCTD Nb, but not wildtype CTD Nb, in transfected 293T cells. Harvested 1 day post-transfection. βgal is a transfection control. Image representative of triplicate transfection sets.

Figure 4.8 (Continued)



dCTD Nb-mCherry, to a similar extent as dGBP1-mCherry, with reduced mCherry fluorescence compared to wildtype and widespread aggregations seen throughout the cells (Figure 4.8B and C). Strikingly, the aggregation was eliminated by co-expression with CTD, suggesting that the CTD/CTD Nb complex becomes stabilized upon complex formation (Figure 4.8B). Western blot analysis further confirmed that CTD stabilizes dCTD Nb-mCherry expression (Figure 4.8E). With mCherry antibody, we also detected the presence of a protein fragment slightly smaller in size than the expected dGBP1-mCherry fusion protein. This could possibly be a degradation product liberating mCherry from dGBP1. We roughly divided the 6 dGBP1 mutations into an N- and C-terminal group (N-terminal: A25V, E63V, S73R. C-terminal: C/S98Y, Q109H, S117F), and tested whether each group could confer CTD Nb with destabilizing phenotypes. Only the C-terminal group gave visibly detectable destabilization of CTD Nb, in the form of aggregations and exclusion from nucleus (Figure 4.8D). This destabilizing phenotype can be rescued by CTD co-expression (Figure 4.8D). The likely key destabilizing residues are C/S98Y and S117F, as we found earlier in a residue mapping experiment of dGBP1 (Figure 4.5). Taken together, our results demonstrate the feasibility of transferring protein-destabilizing mutations to different nanobodies that target different antigens, for rapid generation of intracellular antigen-inducible control.

Discussion

The studies reported here have provided several advances in the development of protein-inducible systems. First, we simplified the use of an intracellular protein with no defined regulatory abilities, GFP, into a driver molecule that can activate an output protein of

interest by interaction with a single responder protein component. This simplification enhanced the efficiency of activation compared to previous dimerizer systems. We also anticipate this system to overcome potential issues from excessive target levels, as seen in dimerizer systems. The system was shown to regulate the activity of fluorescent proteins and site-specific recombinases, demonstrating the utility of such an approach for generating protein-responsive sensors and effectors. The tools developed here should be immediately applicable for studies in the mouse by making use of existing transgenic GFP reporter lines for cell-specific manipulation studies. Notably, it is now simple to perform intersectional genetics between GFP and Flp or Cre driver lines, by using GFP to directly turn on Cre or Flpo recombination. Second, we demonstrated that mutagenesis screens could be employed to discover protein mutations that confer antigen-induced stabilization effects. This paves the way for screening of other nanobodies and/or protein scaffold binders to generate additional destabilizing systems controlled by intracellular proteins. Third, we showed that it is possible to graft destabilizing mutations from one conserved protein scaffold or antibody framework to another, regardless of antigen identity. If further development improves the success rate of conferring antigen-stabilized effects, this strategy could allow for rapid generation of protein sensors and responders without going through laborious screens to discover relevant mutations. Ultimately, the ability to turn any intracellular protein target into a synthetic switch would allow the use of wildtype animals for cell-specific manipulation studies.

Despite having virtually no background signal associated with the dGBP1-TagBFP construct, fusion of dGBP1 to other proteins has variable success. Fusion to mCherry led

to aggregation of the fusion protein in cells, and dGBP1-Cre fusion still display a significant number of background recombination events. Additional engineering efforts will be needed to address the background noise issue with some fusion constructs, but the possible solutions are obvious and under investigations. For example, it would be desirable to incorporate temporal control on the activity of dGBP1-Cre fusion protein by adding an ERT2 domain for sequestering the protein away from the nucleus until introduction of the Tamoxifen ligand. Preliminary data suggest that this strategy works (data not shown).

Potential application of destabilized nanobodies beyond cell-specific manipulations:

To reduce toxic effects with previous GFP-dependent systems

One potential application of dGBPs is to replace the wildtype GBP domain used in the previous T-DDOG and CRE-DOG systems. In the transcription system, overexpression of transactivation domain can lead to squelching of transcription machinery and thereby cell toxicity (Gill and Ptashne, 1988). The negative effects of transcription activation domains may be minimized by suppression of their expression until a cell expresses GFP. In addition, the background recombination seen with GFP-independent association of split Cre components may be further reduced by destabilizing the components in the absence of GFP.

As an improved protein localization probe

Besides being used as reagents to generate protein responsive effectors, destabilized nanobodies could also be exploited as improved probes for protein localization. By

fusing nanobodies to fluorophores, one can visualize the localization of target antigens in living cells (Rothbauer et al., 2006). However, a caveat of this approach is that good signal-to-noise detection of target protein localization requires the nanobody-fluorophore be restricted in localization to the site of interactions. Recently, this issue was addressed by deigning a synthetic circuit with large fusion intrabody protein constructs and transcriptional feedback mechanism (Gross et al., 2013). However, this method requires that the targeted protein be excluded from the nucleus while bound to the antigen, and so has been only demonstrated for proteins anchored away from the nucleus. In comparison, destabilized nanobody-fluorophores would not be limited to extra-nuclear proteins as it is based on destabilization without antigen. We expect that under ideal conditions, any unbound fluorophore would be sent for degradation, effectively suppressing background noise.

Mechanism of destabilization

Proteins are degraded through the ubiquitin pathway, the lysosome, or phagocytosis. We hypothesize that the ubiquitin pathway would mediate the destabilization effects seen in our system. Testing this idea would involve use of ubiquitin ligase blockers or other blockers that inhibit the specific branches of the degradation pathway. Pulse-chase experiments and subcellular-localization analysis of nanobody-halo fusion proteins with or without blocker should address its mechanism of removal.

A caveat of the degradation approach compared to transcription activation of a gene is that the functional protein product is produced and could potentially exert its functional

effects before being sent for degradation. The different amount of background fluorescence we saw with different fluorescent proteins as fusion partners may be due to differences in the rate at which different dGBP1-fusion partner constructs were sent to the proteasome and degraded. One way to address this caveat is to enhance the efficiency of degradation by increasing the number of destabilized nanobodies fused to the fusion partner. We showed that this approach indeed works well to suppress leakage of Cre and especially Flpo activity (Figure 4). Screening for additional destabilizing mutant combinations or further mutagenesis of dGBP1 fused to highly stable fusion partners may further enhance the degradation efficiency.

Reason for lack of success in grafting dGBP1 mutations across nanobodies?

The major contributing GBP1 residues involved in destabilization have been mapped to S73, C98 and S117. All residues fall in the relatively conserved framework regions of the nanobody scaffold. S73 is exposed to the hydrophilic environment. C98 is involved in disulfide bond formation in the scaffold, but disulfide bonds may not be formed in the reducing intracellular environment. S117 is on the very end of the nanobody. In the case of HIV CTD Nb, sub-mapping experiment further suggest that C98 and S117 might be sufficient to confer antigen-inducible stabilization. This assumes that the relative destabilizing contributions of Q109H are also negligible when transferred across nanobodies. Why were these mutations not effective with the other nanobodies tested? The issue seems not with destabilization, since all nanobodies receiving the dGBP1 mutations seemed to be destabilized (data not shown). Rather, what failed in almost all the tests was the ability of the antigen to reverse the destabilized phenotype upon

nanobody binding. A key difference between the antigen binding mode of GBP1, CTD Nb and the other nanobodies chosen in our work is the existence of an elongated loop structure in the CDR3 region that pushes these other antigens away from FR2 and CDR2 of the other nanobodies. The antigen probably acts like a staple in the case of GBP1 and CTD Nb to stabilize any destabilizing effects conferred by C98 and S117. When the antigen is bound primarily to the CDR3 loop, however, the antigen binds away from the backbone structure and so there is no stapling effect, resulting in continued destabilization of the entire structure.

I thus hypothesize that the mode of antigen-nanobody binding will affect the likelihood that grafting of specific mutations between nanobodies will succeed. This idea is supported by the fact that the lone success in grafting dGBP1 mutations was with the CTD-CTD Nb complex, which exhibited a similar mode of binding as GFP-GBP1. There are other reasons why the grafting experiments may have failed for the other nanobodies, such as relatively weaker affinity between selected antigens and nanobodies, defective binding between antigens and nanobodies under assay conditions and inadequate antigen expression. Follow up studies will be needed to exclude these reasons for the grafting failures. Nevertheless, a prediction of my hypothesis is that nanobodies that exhibit more similar modes of interaction with their antigens should be more receptive to transfer of destabilizing mutations with each other, than with nanobodies which bind their antigens in more dissimilar ways. To test such an idea it will be important to establish objective criteria for classifying antigen-nanobody binding modes, to repeat the mutagenesis screens on selected candidates within each clade of complexes,

and then to compare and contrast isolated mutations within clade and across clades. Upon arrival at a common set of residues and residue mutations within and between clades, single and/or combinations of mutations would be grafted between nanobodies to assess whether a clade-specific or universal mutation code can be arrived at for generating antigen-inducible protein stabilization.

Materials and Methods

Animals – All animal experiments performed were approved by the Institutional Animal Care and Use Committee at Harvard University. Time pregnant CD1 (Charles River Breeding Laboratories, Boston, MA) were used for the electroporation experiment.

Generation of mutagenized GBP1 Library:

GBP1 and vector flanking sequences were PCR amplified from the pBMN-GBP1-TagBFP vector. The amplified product was gel extracted and randomly mutagenized using primers corresponding to GBP flanking sequences. The GeneMorph II Random Mutagenesis Kit was used to introduce balanced mutation rates for different nucleotides (Agilent). Duplicates of low, medium, and high mutation reactions were cloned into the pBMN vector by Gibson Assembly. The pBMN backbone was digested with BamHI and SphI to receive the inserts. Next, Gibson reactions were transformed into Stbl-2 cells. Transformants were grown in 200 mL LB cultures overnight, and DNA was purified by Maxiprep kits (QIAGEN). An aliquot of cultures that grew well were plated, and sequenced. The low mutation library had an A→T mutation in the first coding ATG in all sequenced colonies and was not used. Mutagenesis library replicate 2 was used to generate VSV-G coated MLV viruses, and stably infect 293 cells.

Selection of candidate GBP mutants:

Stably infected 293 cells were sorted by FACS for either low, medium, or high red fluorescence (from the IRES HcRed-t) and for low blue fluorescence. Next, cells were plated and allowed expand in culture. Cells were then infected with an 2/8 AAV-EF1a-GFP virus, and allowed to grow for 24 hours. Cells were then sorted for red fluorescence and either high or very high blue fluorescence. Sorted cells were then seeded into T25 flasks and allowed to grow to confluence. Next, DNA from cells was extracted using the DNeasy kit (QIAGEN). Next, GBP sequences were PCR amplified using flanking GBP sequences (above) and Phusion polymerase (New England Biolabs). PCR products were gel purified and cloned back into the pBMN vector using Gibson Assembly. Next, the ligation was PCR purified, transformed, grown for 1 hour in SOC at 37C and seeded onto Carbenicillin plates. 200 colonies were randomly picked and sequenced. Plasmids from clones were individually transfected into 96 well plates of 293T cells with or without CAG-YFP to assay for TagBFP fluorescence.

Molecular Biology – DNA constructs were generated with standard techniques. GBP1 was acquired previously (3K1K, RCSB Protein Data Bank). (Tang et al., 2013).

Miscellaneous Plasmids – pCAG-GFP (Addgene plasmid 11150) (Matsuda and Cepko, 2004), pCAG-YFP (Addgene plasmid 11180) (Matsuda and Cepko, 2004), pRho-GFP-IRES-AP (referred to as Rho-GFP in main text) (Emerson and Cepko, 2011). pCAG-nlacZ (Cepko lab, Harvard Medical School), pCAGEN (empty vector).

pBMN-GBP1-TagBFP - A GBP1-TagBFP construct inserted into a BamHI/NotI digested pBMN-DHFR(DD)-YFP vector (Addgene plasmid 29325), replacing the

DHFR(DD)-YFP insert and generating pBMN-GBP1-TagBFP vector. This becomes the host vector for mutagenized GBP1 inserts.

pBMN-dGBP1-Cre and pBMN-dGBP1-Flpo - pBMN-dGBP1-TagBFP was digested with SphI/SalI, liberating TagBFP as well as the IRES-HcRed-t element. PCR-amplified Cre and Flpo fragments were then inserted into the digested vector via Gibson Assembly.

pBMN-dGBP1x2-Cre and pBMN-dGBP1x2-Flpo – pBMN-dGBP1-Cre or –Flpo plasmids were digested with SphI. gBlock fragment encoding a codon modified dGBP1 was inserted into this site via Gibson Assembly, generating pBMN-dGBP1x2-Cre or –Flpo.

pCAG-CTD – a gblock fragment carrying the CTD gene was inserted into EcoRI/NotI digested pCAG-GFP via Gibson Assembly; CTD replaces GFP in the cassette.

pCAG-dGBP1-mCherry – PCR amplified mCherry was inserted into a SphI/NotI digested pCAG-dGBP1-TagBFP vector, resulting in replacement of TagBFP with mCherry. The vector becomes pCAG-dGBP1-mCherry.

Wildtype and mutant nanobodies in pCAG vector – all variant sequences were synthesized as gblocks and inserted into EcoRI/SphI digested pCAG-dGBP1-mCherry or pCAG-dGBP1-TagBFP via Gibson Assembly. The insert fragment thus replaces dGBP1.

Cell culture and Transfection. 293T cells were seeded onto 24 or 96 well plates and used for transfection when the cells reach between 60-95% confluency, usually 1-2 days later. Transfection is achieved with polyethyleneimine (PEI) at a 1:4 DNA amount:PEI volume ratio. Between 100 and 400 ng total DNA were transfected into single wells of

96 well plates for fluorescence analysis of destabilized mutants. Around 400-500 ng total DNA were transfected into single wells of 24 well plates for western blot analysis.

Western Blot: 293T cells were seeded onto 24 well plate and transfected with the relevant constructs plus a constant amount of CAG-nlacZ plasmid, which serves as a transfection control. Transfected 293T cells were lysed in 6x SDS PAGE loading buffer (350mM Tris-HCl, pH8, 30% glycerol, 10% SDS, 600mM DTT, 0.01% Bromophenol Blue) and stored at -20 degrees Celsius until used for SDS PAGE and Western blot.

***In vivo* electroporations –** P0 CD1 retinas were microinjected with plasmids into the subretinal space and subjected to electroporation (Matsuda and Cepko, 2004). Plasmid encoding CAG-dGBP1-TagBFP was injected along with plasmids CAG-DsRed, CAG-GFP or Rho-GFP. Electroporated retinas were harvested at P14, immunostained for anti-TagBFP in the far red channel, and imaged by confocal microscopy.

Retinal Histology: Isolated mouse retinas were fixed at room temperature for 30 minutes in 4% paraformaldehyde/PBS solution. Retinas were then transferred to 30% sucrose in PBS, and subsequently into a 50/50 mixture of 30% sucrose/PBS and OCT for sectioning. 20 μm retinal cryosections were cut on a Leica CM3050 cryostat (Leica Microsystems), using disposable blades.

Retinal immunohistochemistry: Retinal cryosections were incubated in blocking solution (3% normal goat serum, 1% BSA, 0.1% Triton-X, 0.02% SDS in PBS) for 1 hour and stained for primary antibody overnight at 4 degrees Celsius. Immunostained cryosections were washed three times in PBS and stained for secondary antibodies in blocking solution for 2 hours at room temperature. Slides were then washed in PBS and

mounted for imaging in Fluoromount-G (Southern Biotechnology Associates; 0100-01).

General Microscopy and Image Analysis. Retinal section images were acquired on a Zeiss LSM780 confocal microscope, on a 40x oil immersion objective. Images were analyzed and processed on Imaris, ImageJ and/or Photoshop software. Cell culture images were acquired on a Leica DMI3000B microscope, using a 5x, 10x or 20x objective.

Antibodies: Antibodies used were rabbit anti-TagRFP (also targets TagBFP; 1:1000 dilution for both immunoblot and immunohistochemistry) (AB233, Evrogen), mouse anti-βgal (1:50 dilution for immunoblot) (40-1a supernatant, Developmental Studies Hybridoma Bank), rabbit anti-DsRed (also targets mCherry; 1:1000 or 1:2000 dilution for immunoblot) (632392, Clontech), rabbit-anti-GFP (1:500 dilution for immunohistochemistry) (A-6455, Invitrogen). Secondary antibodies raised against the appropriate species were acquired from Jackson ImmunoResearch or Invitrogen.

Statistical Analysis - Two-tailed Student's t test assuming unequal variance was used for all comparisons. $p > 0.05$ is judged as statistically significant.

References. Osamu Shimomura - Nobel Lecture: Discovery of Green Fluorescent Protein, GFP" (Nobel Media AB 2014).

Banaszynski, L.A., Chen, L.C., Maynard-Smith, L.A., Ooi, A.G., and Wandless, T.J. (2006). A rapid, reversible, and tunable method to regulate protein function in living cells using synthetic small molecules. Cell *126*, 995-1004.

Bonger, K.M., Rakhit, R., Payumo, A.Y., Chen, J.K., and Wandless, T.J. (2014). General method for regulating protein stability with light. ACS Chem Biol *9*, 111-115.

Brand, A.H., and Perrimon, N. (1993). Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. Development *118*, 401-415.

Butala, M., Zgur-Bertok, D., and Busby, S.J. (2009). The bacterial LexA transcriptional repressor. Cell Mol Life Sci *66*, 82-93.

Dymecki, S.M. (1996). Flp recombinase promotes site-specific DNA recombination in embryonic stem cells and transgenic mice. Proc Natl Acad Sci U S A *93*, 6191-6196.

Emerson, M.M., and Cepko, C.L. (2011). Identification of a retina-specific Otx2 enhancer element active in immature developing photoreceptors. Dev Biol *360*, 241-255.

Gill, G., and Ptashne, M. (1988). Negative effect of the transcriptional activator GAL4. Nature *334*, 721-724.

Gross, G.G., Junge, J.A., Mora, R.J., Kwon, H.B., Olson, C.A., Takahashi, T.T., Liman, E.R., Ellis-Davies, G.C., McGee, A.W., Sabatini, B.L.*, et al.* (2013). Recombinant probes for visualizing endogenous synaptic proteins in living neurons. Neuron *78*, 971-985.

Gurskaya, N.G., Fradkov, A.F., Terskikh, A., Matz, M.V., Labas, Y.A., Martynov, V.I., Yanushevich, Y.G., Lukyanov, K.A., and Lukyanov, S.A. (2001). GFP-like chromoproteins as a source of far-red fluorescent proteins. FEBS Lett *507*, 16-20.

Kirchhofer, A., Helma, J., Schmidthals, K., Frauer, C., Cui, S., Karcher, A., Pellis, M., Muyldermans, S., Casas-Delucchi, C.S., Cardoso, M.C.*, et al.* (2010). Modulation of protein properties in living cells using nanobodies. Nat Struct Mol Biol *17*, 133-138.

Matsuda, T., and Cepko, C.L. (2004). Electroporation and RNA interference in the rodent retina in vivo and in vitro. Proc Natl Acad Sci U S A *101*, 16-22.

Muyldermans, S. (2013). Nanobodies: natural single-domain antibodies. Annu Rev Biochem *82*, 775-797.

Orban, P.C., Chui, D., and Marth, J.D. (1992). Tissue- and site-specific DNA recombination in transgenic mice. Proc Natl Acad Sci U S A *89*, 6861-6865.

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

Rothbauer, U., Zolghadr, K., Tillib, S., Nowak, D., Schermelleh, L., Gahl, A., Backmann, N., Conrath, K., Muyldermans, S., Cardoso, M.C*., et al.* (2006). Targeting and tracing antigens in live cells with fluorescent nanobodies. Nat Methods *3*, 887-889.

Schonig, K., Bujard, H., and Gossen, M. (2010). The power of reversibility regulating gene activities via tetracycline-controlled transcription. Methods Enzymol *477*, 429-453.

Subach, O.M., Gundorov, I.S., Yoshimura, M., Subach, F.V., Zhang, J., Gruenwald, D., Souslova, E.A., Chudakov, D.M., and Verkhusha, V.V. (2008). Conversion of red fluorescent protein into a bright blue probe. Chem Biol *15*, 1116-1124.

Tang, J.C., Szikra, T., Kozorovitskiy, Y., Teixiera, M., Sabatini, B.L., Roska, B., and Cepko, C.L. (2013). A nanobody-based system using fluorescent proteins as scaffolds for cell-specific gene manipulation. Cell *154*, 928-939.

Chapter Five

Concluding Discussion

Fluorescent proteins as multifunctional switches, and protein-responsive systems

In this thesis, I focused on the issue of how to expand the utility of the many transgenic GFP reporter lines generated over the past 2 decades. My work shows that GFP, its derivatives, and other more distantly related fluorescent proteins could potentially all be exploited as signals regulating synthetic systems. I demonstrate here that fluorescent proteins can be thought of as scaffolds to bring split protein components together (Chapter 2 and 3) and as protein stabilizers to prevent misfolding and/or degradation of output proteins (Chapter 4) (Figure 5.1). Since fluorescent proteins were originally applied for their fluorescent abilities, the potential use of fluorescent proteins as signaling molecules remain underexplored. In the near future it is likely that GFP would be exploited as a transmembrane signaling ligand analogous to Notch signaling (Ehebauer et al., 2006; Struhl and Adachi, 1998) or as a diffusible ligand bringing together split membrane receptors (Wehr et al., 2006).

The significance of my work lies beyond fluorescent proteins, as the multiple synthetic strategies employed here to co-opt GFP should also be useful to exploit other intracellular products, RNA or proteins, for cell-specific gene manipulation. This has the potential to extend the biologist's ability to experiment with wildtype animals, as one simply has to target an intracellular product for cell-specific control, without having to undertake the more laborious approach of discovering cis-regulatory elements underlying the gene's expression pattern, and/or the need to modify endogenous gene loci with transgene knockin. In the context of studies using transient gene delivery tools like viruses and electroporation to target cell types, the approach taken here should prove more efficient

than genome editing approach, which currently suffers from low transgene integration efficiency in postmitotic cells (personal communications, Sui Wang). Nevertheless, I envision the combination of genome editing and intracellular product targeting will complement each other in the future, as it should be possible to integrate the effects of knockin transgenes from multiple loci at the protein-protein interaction level to generate increasingly precise intersectional gene control.

The bottlenecks to extending my approaches to any intracellular product of interest are the availability of reagents for binding desired target, a simple and generalizable approach converting the binding event into a desired molecular output, and a rapid way of generating protein responders to any intracellular targets. These are being addressed here and elsewhere. First, much effort is being invested to isolate artificially derived binding proteins (Wurch et al., 2012). Second, I introduced a one-component strategy whereby an input intracellular protein can control the activity of a variety of desired output proteins, simply by protein stabilization (Chapter 4). Third, I showed that it is possible to rapidly engineer sensors and effectors responsive to different intracellular proteins, simply by grafting mutations across a conserved protein scaffold or antibody framework to confer antigen-dependent stabilization (Chapter 4). I anticipate current studies in our laboratory to further establish the broad applicability of the protein destabilization approach in the near future.

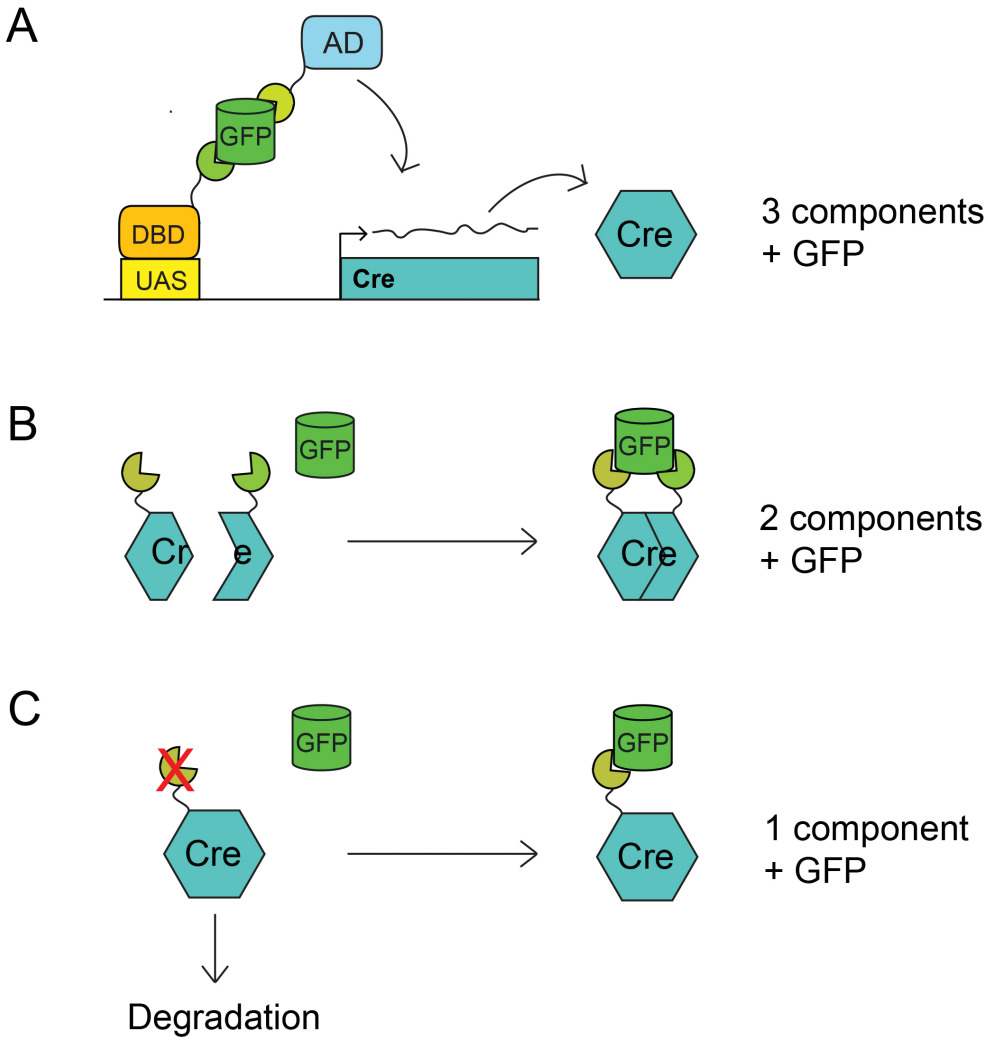


Figure 5.1. Summary of GFP-dependent systems developed in this thesis. A. GFP-dependent transcription system described in Chapter 2. B. GFP-dependent Cre recombinase described in Chapter 3. C. GFP-dependent protein stabilization system described in Chapter 4.

Natural function of fluorescent proteins and other molecular tools

Over the past few decades, much effort has been placed on mining the natural world for gene products with useful biological properties. These efforts have resulted in the expansion of the repertoire of fluorescent proteins as well as additional functional molecular tools such as transcription regulators, light-sensing ion channels and pumps, and genome editing systems (Cong et al., 2013; Luo et al., 2008; Shaner et al., 2005; Yizhar et al., 2011). However, given that many useful gene products are now being applied in heterologous systems, one could consider expanding the uses of these tools beyond their original applications. To illustrate, fluorescent proteins were originally chosen and promoted as reporter molecules for ability to fluoresce without the need for cofactor. This drove the direction of fluorescent protein development towards imaging based applications, resulting in many useful spin-off uses such as calcium imaging and synaptic release imaging (Miesenbock et al., 1998; Nakai et al., 2001). However, the utility of fluorescent proteins as demonstrated here is not necessarily confined to their fluorescent ability. The fact that GFP is expressed in different cell populations amongst transgenic lines allows it to be used as a cell-specific driver molecule, by interfacing it with synthetic systems. By extending this logic to other molecular tools, it becomes interesting to consider whether channelrhodopsins, for example, could be utilized for other light-independent uses, perhaps such as artificial signaling systems if it was found that ChR2 can be modulated in its host by certain interaction partners. Further, to further explore the interactome of Cas9 proteins may reveal novel uses of the CRISPR system beyond genome editing or genomic loci imaging/manipulation. After all, binary systems and more complex systems such as the Gal4/UAS, Cre/loxP and Gal4/Gal80 systems, did

require understanding of the biological process and components involved in the native host.

The ability to use GFP as a scaffold protein raises the question of what exactly is fluorescent proteins' endogenous function? Conventional wisdom suggests that fluorescent proteins are useful in marine organisms for their light absorbing and/or emitting properties (Haddock et al., 2010), but this could be a product of our own biased thinking. The fact that fluorescent protein is a protein means that it may interacts physically with other proteins/molecules in their host cells. Indeed, it has been suggested that fluorescent proteins diverged from a class of extracellular matrix proteins called nidogens. The G2 domain of Nidogen-1 shares incredible similarity to fluorescent proteins in their beta barrel structure and their crystal structure alignment shows a very small deviation in the RMS parameter (Hopf et al., 2001). Since Nidogens are not fluorescent, the obvious approach to determine their natural function is to analyze protein-protein interactions. Ironically, Nidogens actually behave like scaffold proteins, interfacing with other extracellular matrix proteins (Hopf et al., 2001). By the same logic, a focus on what fluorescent proteins interact physically with in their host cells could provide novel insight into their biology as well as previously unimaginable applications. We currently know almost nothing about what fluorescent proteins interact with, and if there are interactions in the host cells, what would be the consequences. Studies suggest that GFP binds weakly to the calcium sensor Aequorin, but it has actually has not been shown that this interaction occurs live in its native environment, the jellyfish. Indeed, when misexpressing a nuclear-localized Aequorin with GFP in live

mammalian cells, I failed to detect any interactions between the two molecules even when they are strongly overexpressed in cell culture, suggesting weak or transient interactions or missing cofactors (data not shown). Thus, current limits in knowledge, and the broad applications of GFP and its derivatives, emphasize the need to return to the jellyfish for further characterizations. I recently initiated efforts to address this issue.

Exploiting endogenous, intracellular products for cell-specific manipulations

As many cell types express unique profiles of RNA and proteins, as well as a unique repertoire of metabolic products, all these molecules could potentially be exploited as signals for synthetic systems designed for cell-specific control. Since artificially derived binding proteins can be selected or designed to bind different antigens, it is not unreasonable to believe that in the coming years the strategies taken to exploit GFP as a cell-specific signal will be extended to target other intracellular products for cell-specific gene manipulation in multicellular organisms (Figure 5.2).

Certainly, a concern with this idea is that endogenous intracellular products have naturally evolved functions and interaction partners. By creating novel interactions one might disrupt the natural functions of these target antigens, leading to unintended side effects. Of course, if the desired manipulation is to detect a diseased cell based on based on expression or modification of a protein, and use the protein responders to kill the cell, then it does not matter whether the targeted cell's protein networks is disrupted. However, in many cases one does want to preserve the function of the endogenous cellular network. Even so, protein-protein interaction networks are not fixed – they must

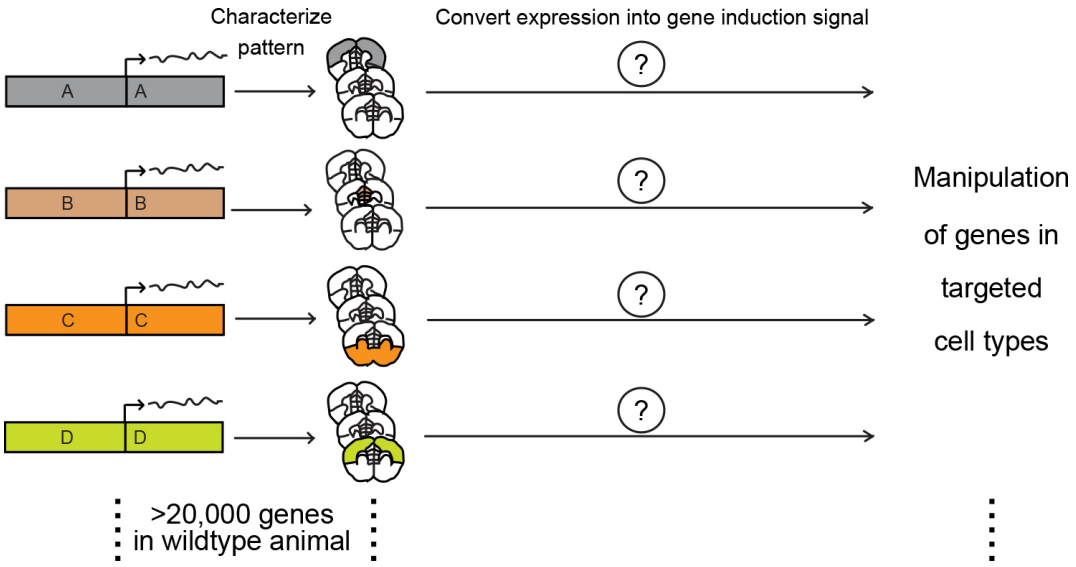


Figure 5.2. Co-opting endogenous intracellular products for cell-specific gene manipulation. In the near future, one can envision using strategies such as to the ones demonstrated in this thesis for targeting and manipulating specific cell types based on expression of their endogenous intracellular products such as RNA and proteins. This approach would bypass the need for transgenic animals expressing a driver molecule in specific cell types.

have some level of variation within the lifetime of a cell, between cells, between genetically variable individuals. Also, the degree of protein connectivity varies amongst proteins, and existing proteins can function via interaction with multiple complexes at different cellular locations. Thus, in actuality endogenous proteins are already exposed to varying cellular states and interaction demands. Whether their current network is rigidly fixed or has flexibility is an interesting issue to consider, because if it has flexibility, then there is space for introducing new artificial connections. One example could be the Mediator complex, Pol II and the ribosome complex – essential core machinery proteins that necessarily have to interact with a whole host of different partners within their lifetime. Thus, it should not be unreasonable to believe that one can artificially assign novel roles for endogenous components, provided we understand the principles behind protein-protein interaction networks. Indeed, it would be fascinating to determine the extent to which novel interactions can be introduced to different protein components in the host cell network without perturbing the original network, and if such manipulations are possible, where would one insert these changes and how.

Summary

In this thesis, I have described synthetic approaches to co-opting an intracellular protein, GFP, for controlling biological activities. First, by exploiting GFP binding nanobodies as tools for constructing synthetic protein complexes, I was able to demonstrate the utility of fluorescent proteins as scaffolds for controlling transcriptional regulation. The hybrid transcription factors described here could be applied *in vivo* to selectively manipulate the various GFP-labeled cell types in transgenic GFP reporter lines for applications such as genetic and optogenetic perturbations.

Second, I explored to use GFP to turn on Cre recombination. A number of optimization screens led to the isolation of a tight GFP-inducible split Cre system whereby GFP can directly activate Cre recombinase activity, bypassing the need for a transcription system to turn on Cre recombinase. I further show that the split Cre system could be made drug-inducible for temporal control. The system can be delivered to the mammalian brain and retina using AAV, easing the way one would manipulate GFP labeled cells in the mouse.

Finally, I was able to develop the simplest possible GFP-dependent system, by using GFP as a signal to control protein stabilization. This binary system overcomes caveat associated with the dimerizer approach, making the activity of the system less sensitive to inhibition by high GFP levels. I again demonstrate the utility of this system in the mouse retina for cell-specific gene manipulation. I show that mutations found in the destabilized GBP1 could be transferred to other nanobodies to confer ligand inducible protein stabilization, suggesting potential to elucidate a universal code for nanobody destabilization and ligand-inducible stabilization.

Together, my work expands our concept of the uses of fluorescent proteins and demonstrates the potential of synthetic approaches for co-opting any intracellular products for cell-specific gene manipulation in multicellular organisms.

References

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A.*, et al.* (2013). Multiplex genome engineering using CRISPR/Cas systems. Science *339*, 819-823.

Ehebauer, M., Hayward, P., and Martinez-Arias, A. (2006). Notch signaling pathway. Sci STKE *2006*, cm7.

Haddock, S.H., Moline, M.A., and Case, J.F. (2010). Bioluminescence in the sea. Ann Rev Mar Sci *2*, 443-493.

Hopf, M., Gohring, W., Ries, A., Timpl, R., and Hohenester, E. (2001). Crystal structure and mutational analysis of a perlecan-binding fragment of nidogen-1. Nat Struct Biol *8*, 634-640.

Luo, L., Callaway, E.M., and Svoboda, K. (2008). Genetic dissection of neural circuits. Neuron *57*, 634-660.

Miesenbock, G., De Angelis, D.A., and Rothman, J.E. (1998). Visualizing secretion and synaptic transmission with pH-sensitive green fluorescent proteins. Nature *394*, 192-195.

Nakai, J., Ohkura, M., and Imoto, K. (2001). A high signal-to-noise Ca(2+) probe composed of a single green fluorescent protein. Nat Biotechnol *19*, 137-141.

Shaner, N.C., Steinbach, P.A., and Tsien, R.Y. (2005). A guide to choosing fluorescent proteins. Nat Methods *2*, 905-909.

Struhl, G., and Adachi, A. (1998). Nuclear access and action of notch in vivo. Cell *93*, 649-660.

Wehr, M.C., Laage, R., Bolz, U., Fischer, T.M., Grunewald, S., Scheek, S., Bach, A., Nave, K.A., and Rossner, M.J. (2006). Monitoring regulated protein-protein interactions using split TEV. Nat Methods *3*, 985-993.

Wurch, T., Pierre, A., and Depil, S. (2012). Novel protein scaffolds as emerging therapeutic proteins: from discovery to clinical proof-of-concept. Trends Biotechnol.

Yizhar, O., Fenno, L.E., Davidson, T.J., Mogri, M., and Deisseroth, K. (2011). Optogenetics in neural systems. Neuron *71*, 9-34.

Appendix I

Supplemental Information

for Chapter 2

Supplemental Information for "A Nanobody-Based System using Fluorescent Proteins as Scaffolds for Cell-Specific Gene Manipulation" by Tang et al. Modified figure references to indicate thesis figures in Chapter 2

EXTENDED EXPERIMENTAL PROCEDURES

Animals

The following mouse strains were used: Timed pregnant CD1 and C57/BL6 mothers (Charles River Breeding Laboratories, Boston, MA), *Otx2*$^{fl/fl}$ mice (Tian et al., 2002). C57/BL6 mice (Charles River Breeding Laboratories or Jackson Laboratories, Bar Harbor, ME). Tg(GUS8.4GFP) (Huang et al., 1999) (Robert Margolskee, Monell Chemical Senses Center, Philadelphia, PA). Zebrafish strains used are Tubingen wild-type and Tg(*ubi-GFP*) zebrafish (Mosimann et al., 2011) (Leonard Zon, Children's hospital, Boston, MA).

Molecular Biology

GBP Sequences

GBP1 and 4 sequences were obtained from published protein sequence (PDB; 3K1K for GBP1 and 3G9A for GBP4), back-translated, codon-optimized for the mouse and synthesized by Genewiz (New Jersey), generating pUC57-GBP1 and pUC57-GBP4. Plasmids carrying GBPs 2, 5, 6 and 7 were obtained from Ulrich Rothbauer (Ludwig Maximilians University Munich). GBP6 was synthesized based on provided sequence from Ulrich Rothbauer to generate pUC57-GBP6 (Genewiz).

Miscellaneous Plasmids

pCAG-GFP (Addgene plasmid 11150) (Matsuda and Cepko, 2004), pCAG-YFP (Addgene plasmid 11180) (Matsuda and Cepko, 2004), pCAG-CFP (Addgene plasmid 11179) (Matsuda and Cepko, 2004), pCAG-tdT (Cepko lab, Harvard Medical School), pCAG-mCherry (Cepko lab, Harvard Medical School), pCAG-dsRed (Addgene plasmid 11151) (Matsuda and Cepko, 2004), pUAS-luc2 (Addgene plasmid 24343) (Potter et al., 2010), pRL-TK (Promega, #E2241), pBS SK+ (Stratagene), mGluR6 (200bp crit reg)+SV40p-GFP-IRES-AP (referred to as mGluR6-GFP in main text) (Addgene plasmid 18817) (Kim et al., 2008), pRho-GFP-IRES-AP (referred to as Rho-GFP in main text) (Emerson and Cepko, 2011), pCAG-nlacZ (Cepko lab, Harvard Medical School).

Plasmid Construction

pCAG-GFPmG1

To produce GFPmG1, splicing by overlap extension (SOE) PCR was performed to introduce E143K and N147Q mutations in EGFP. The mutagenized PCR product has AgeI-Kozak consensus sequence and NotI on the 5′ and 3′ ends, respectively. Using AgeI/NotI, this fragment was cloned in place of EGFP in the pCAG-GFP vector.

pUAS-tdT

TdT was amplified from pCAG-tdT with EcoRI and XbaI restriction sites added on the 5′ and 3′ ends, respectively. This fragment was cloned into pUAS-luc2 (5xUAS) via EcoRI/XbaI, replacing the luciferase sequence to give pUAS-tdT.

pTRE-tdT and pTRE-luc2

TREtight promoter was amplified from pTRE-TIGHT miR-1 (Addgene plasmid 14896) (Lewis et al., 2003), generating SphI and SbfI restriction sites on the 5′end. This fragment was cloned into pUAS-tdT via SphI/EcoRI restriction sites, replacing the UAS-hsp70 sequence and giving pTRE-tdT. The same fragment was digested with SbfI/EcoRI and cloned into the corresponding sites in pUAS-luc2, replacing the UAS-hsp70 sequence and giving pTRE-luc2.

plexAop2-luc2

lexAop2-hsp70 minimal promoter was amplified from pJFRC18-8XLexAop2-mCD8::GFP (Addgene plasmid 26225) (Pfeiffer et al., 2010), generating the EcoRI restriction site on the 3′ end. The PCR fragment was digested with SbfI/EcoRI and cloned into the corresponding sites in pUAS-luc2, replacing UAS-hsp70 sequence and giving pLexAop2-luc2.

pUAS-Cre

Cre recombinase was amplified from pNrl-Cre (Addgene plasmid 13780) (Matsuda and Cepko, 2007), generating an EcoRI-Kozak sequence and XbaI restriction site on the 5′ and 3′ ends, respectively. This fragment was inserted in place of luc2 in the pUAS-luc2 vector via EcoRI/XbaI restriction sites.

p10xUAS-ChR2-mCherry

ChR2/H134R-mCherry amplicon was flanked with EcoRI/XbaI on the 5′ and 3′ ends, respectively, and inserted into a 10xUAS vector via EcoRI/XbaI sites.

pCAG-N-Gal4DBD

The Gal4 DNA-binding domain (Gal4DBD) was amplified from pAcPL-Gal4DBD (Addgene plasmid 15304) (Luan et al., 2006), with AgeI-Kozak consensus sequence and NheI-10Glycine-MfeI-NotI overhangs on the 5′ and 3′ ends, respectively. This fragment was inserted in place of EGFP in the pCAG-GFP vector via AgeI/NotI restriction sites.

pCAG-C-Gal4DBD

The GAL4 DNA-binding domain (Gal4DBD) was amplified from pAcPL-Gal4DBD (Addgene plasmid 15304) (Luan et al., 2006), with AgeI-NheI and NotI overhangs added on the 5′ and 3′ ends, respectively. This fragment was inserted in place of EGFP in the pCAG-GFP vector via AgeI/NotI restriction sites.

pCAG-VP16AD

The VP16 activation domain (VP16AD), along with a NheI-10Glycine polylinker on the 3′ end, was amplified from pAcPL-VP16 (Addgene plasmid 15305) (Luan et al., 2006), generating AgeI-Kozak consensus sequence and NotI restriction site on the 5′ and 3′ ends, respectively. This fragment was inserted in place of EGFP in the pCAG-GFP vector via AgeI/NotI restriction sites.

GBP1+6 T-DDOG Constructs

GBP1-Containing Plasmids

In all GBP1-containing constructs, GBP1 was amplified from pUC57-GBP1 with primers bearing various overhangs on the PCR products.

pCAG-GBP1-10 gly-Gal4DBD. AgeI-Kozak-GBP1-NheI PCR fragment was cloned into pCAG-C-Gal4DBD via AgeI/NheI sites.

pCAG-Gal4DBD-10 gly-GBP1. NheI-10 gly-MfeI-GBP1-NotI PCR fragment was cloned into pCAG-N-Gal4DBD via NheI/NotI sites.

pCAG-Gal4DBD-GBP1. NheI-GBP1-NotI PCR fragment was cloned into pCAG-N-Gal4DBD via NheI/NotI sites.

pCAG-Gal4DBD-GBP1x2. NheI-GBP1-NheI PCR fragment was cloned into pCAG-Gal4DBD-10 gly-GBP1 via NheI restriction site.

pCAG-VP16AD-10 gly-GBP1. NheI-10 gly-MfeI-GBP1-NotI PCR fragment was cloned into pCAG-VP16AD via NheI/NotI sites.

pCAG-rTetRDBD-GBP1. The DNA-binding domain of Reverse Tetracycline transactivator 3G (rtTA3G) was amplified from pLenti CMV rtTA3G Blast (R980-M38-658) (Addgene plasmid 31797, Dominic Esposito). The PCR product contains AgeI-Kozak consensus sequence and NheI restriction site on the 5′ and 3′ ends, respectively. This fragment was cloned into pCAG-Gal4DBD-GBP1 via AgeI/NheI and replaces Gal4DBD.

pCAG-GBP1-10 gly-lexADBD. The LexA DNA-binding domain (LexA DBD) was amplified from pCMV Lex VP16 HA (P#1708) (Addgene plasmid 14593) (Kurosu and Peterlin, 2004) with NheI-10 gly-XbaI and NotI overhangs on the 5′ and 3′ ends, respectively. This fragment was cloned into pCAG-GBP1-10 gly-Gal4DBD via NheI/NotI and replaces Gal4DBD.

GBP6-Containing Plasmids

For all GBP6-containing constructs, GBP6 was amplified from pUC57-GBP6 with various overhangs on the PCR products.

pCAG-Gal4DBD-GBP6. NheI-GBP6-NotI PCR product was cloned into pCAG-N-Gal4DBD via NheI/NotI restriction sites.

pCAG-VP16AD-GBP6. NheI-GBP6-NotI PCR product was cloned into pCAG-VP16AD via NheI/NotI restriction sites

pCAG-GBP6-10 gly-Gal4DBD. AgeI-Kozak consensus-GBP6-NheI was cloned into pCAG-C-Gal4DBD via AgeI/NheI sites.

pCAG-GBP6-10 gly-VPminx2, pCAG-GBP6-10 gly-VPminx3, pCAG-GBP6-10 gly-VPminx4. The VPminx2, x3, x4 sequences were amplified from CMV rtTA3G Blast (R980-M38-658) (Addgene plasmid 31797, Dominic Esposito) with NheI-10 gly-MfeI and NotI overhangs on the 5′ and 3′ end, respectively. Each PCR fragment was separately cloned into pCAG-GBP6-10 gly-Gal4DBD.

pCAG-p65AD-GBP6. The p65 activation domain (p65AD) was amplified from pCMV4 p65 (Addgene plasmid 21966) (Ballard et al., 1992) with the NLS from SV40 large T-antigen. This fragment was amplified again to add AgeI-Kozak consensus sequence and NheI on the 5′ and 3′ end, respectively. The final PCR product was cloned into pCAG-VP16AD-GBP6 via AgeI/NheI, replacing VP16AD.

IRES-Linked, GBP1+6 T-DDOG Constructs

pCS2+-Gal4DBD-GBP1-IRES-p65AD-GBP6. Gal4DBD–GBP1 and p65AD-GBP6 were linked with an IRES by SOE PCR, adding BamHI/EcoRI on the 5′ and 3′ ends, respectively. This fusion PCR product was cloned into pCS2+ via BamHI/EcoRI sites.

GBP2+7 T-DDOG Constructs

pCAG-GBP2-10 gly-Gal4DBD. GBP2 was amplified from GBP2 chromobody plasmid with AgeI-Kozak and NheI sequences on the 5′ and 3′ ends, respectively. This fragment was cloned into pCAG-C-Gal4DBD via AgeI/NheI sites.

pCAG-VP16AD-10 gly-GBP7. GBP7 was amplified from GBP7 chromobody plasmid with NheI-10 gly-MfeI on the 5′ end and NotI on the 3′ end. This fragment was cloned into pCAG-VP16AD via NheI/NotI sites.

pCAG-Gal4DBD-GBP2. GBP2 was amplified from GBP2 chromobody plasmid with NheI and NotI sequences on the 5′ and 3′ ends, respectively. The PCR product was cloned into pCAG-N-Gal4DBD via NheI/NotI restriction sites.

pCAG-GBP7-p65AD. GBP7 was amplified from pCAG-p65AD-10 gly-GBP7, adding AgeI-Kozak-NLS and MfeI sequences on the 5′ and 3′ ends, respectively. The PCR product was cloned into pCAG-GBP7-10 gly-p65AD via AgeI/MfeI, replacing NLS-less GBP7 and a 10 glycine linker.

GBP1+5 T-DDOG Constructs

pCAG-GBP5-10 gly-Gal4DBD. GBP5 was amplified from GBP5 chromobody plasmid with AgeI-Kozak consensus sequences and NheI on the 5′ and 3′ end, respectively. The final PCR product was inserted into pCAG-C-Gal4DBD via AgeI/NheI, replacing GBP1.

pCAG-GBP1-10 gly-VPminx2. VPminx2 sequences were amplified from CMV rtTA3G Blast (R980-M38-658) (Addgene plasmid 31797, Dominic Esposito) with NheI-10 gly-MfeI and NotI overhangs on the 5′ and 3′ end, respectively. The PCR fragment was cloned into pCAG-GBP1-10 gly-Gal4DBD via NheI/NotI, replacing 10 gly-Gal4DBD.

Cell Culture and Transfection

Unless stated otherwise, for all cell culture experiments, 1-2x10^5 293T cells were seeded into 48 well plates and 1-2 days later transfected with plasmids. Plasmids were transfected via polyethyleneimine (PEI) at a 1:4 DNA amount:PEI volume ratio. For doxycycline-inducible experiments, doxycycline hyclate (Sigma, D9891-10G) was diluted in water and used at concentrations up to 1 μg/ml.

Cell Culture tdT Reporter Readout

A total of 500 ng of DNA were transfected. In all experiments, 100 ng of UAS-tdT or TRE-tdT were included. Plasmids encoding CAG-driven XFP, DBDG, ADG and other variants were transfected at amounts adjusted for their molarity. pBS SK+, pCAGEN, or

pCAG-mCherry were added to adjust the total DNA amount to equal levels. Fluorescent micrographs were taken on a Leica DMI3000B microscope with a 10x or 20x objective.

Luciferase Assay

In all experiments, 12.5 ng pUAS-luc2 and 1.25 ng pRL-TK were included. Plasmids encoding CAG-driven XFP, DBDG, ADG and other variants were transfected at amounts adjusted for their molarity. pBS SK+ or pCAGEN was added to adjust the total DNA amount to 62.5 or 63.5 ng. Cells were harvested 24 hr later for dual-luciferase assay (Promega) according to manufacturer's instructions. Lysates were pipetted into 96-well plates and read in an Analyst GT plate reader (Molecular Devices). To determine the linear range of detection for the plate reader, a standard curve was constructed by measuring luciferase activity of serial dilutions of QuantiLum recombinant luciferase (Promega). Transfection amounts were then optimized to give readings within the linear range of detection for the instrument. Comparisons between experimental conditions were only done when cell lysates were assayed at the same time. All transfection conditions were repeated 2 or 3 times. Cell lysate from each repeat was split into 3 samples for luciferase assay. 2-tailed Student t test assuming unequal variance was used to test for significance, judged to be $p < 0.05$.

Dosage Response Curve Experiment

A total of 100 ng total DNA were transfected. 15.5 ng of pCAG-Gal4DBD-GBP6 and pCAG-VP16-10lk-GBP1, 12.5 ng pUAS-luc2 and 1.25 ng pRL-TK were included. CAG-GFP plasmid was serially diluted 3 fold in water and pipetted at equal volume into transfection mixture. pCAG-mCherry plasmid was used to make up the total DNA amount. 24 hr after transfection, cells were imaged on Leica epifluorescence microscope at 20x for GFP and mCherry fluorescence before being used for dual-luciferase assay (Promega). All transfection conditions were repeated 3 times. Cell lysate from each repeat was split into 3 samples for luciferase assay.

General Microscopy and Image Analysis

Retinal sections or whole-mount retinas were taken on a Zeiss LSM780 confocal microscope. Slides were scanned using a 40x or 63x oil immersion objective. Cell culture and zebrafish images were images were obtained on a Leica DMI3000B epifluorescence microscope. Whenever possible, image settings were adjusted for saturation. Whenever samples were to be compared within an experiment, image settings and processing were kept constant. Images were analyzed and processed on Imaris, Image J and/or Photoshop software. Images from in vivo electroporation were smoothened on Imaris using the median filter as 3x3x1pixel dimension. Image contrast was adjusted either in Imaris and/or Photoshop.

Retinal Histology

Mouse retinas were dissected out of the eyes or harvested from culture and fixed at room temperature for 30 min in 4% paraformaldehyde. Fixed retinas were washed in PBS and equilibrated in increasing concentration of sucrose (10/30%) 1xPBS pH 7.4 solutions. Retinas were then equilibrated in OCT for at least 10 min and quickly frozen on dry ice. Retinal cryosections were cut into 20 or 25 μm slices on a Leica CM3050S cryostat (Leica Microsystems), using disposable blades.

Retinal Immunohistochemistry

Retinal cryosections were blocked in 5% heat-inactivated normal goat serum in 0.1% TritonX-1xPBS (PBT), pH7.4 for 1 hr at room temperature and then stained with primary antibody in blocking solution overnight at 4°C. Slides were washed in 1xPBT for three times and then incubated in secondary antibodies and DAPI (0.3 μM) for 2 hr at room temperature. Slides were then washed in 1xPBT for three times and mounted using Fluormount-G (Southern Biotechnology Associates; 0100-01).

Antibodies

Antibodies used in this study were: chicken anti-β-galactosidase (1:1000 dilution) (ab9361, Abcam), rabbit anti-OTX2 (1:500 dilution) (AB9566, Chemicon), mouse anti-PAX6 (1:20 dilution) (Pax6 s, Developmental Studies Hybridoma Bank), goat anti-ChAT (1:100 dilution for retinal section, 1:200 dilution for wholemount retina) (AB144, Millipore), rabbit anti-GFP (1:500 dilution) (A-6455, Invitrogen) or chicken anti-GFP (1:1000 dilution) (ab13970) were only used on Tg(GUS8.4-GFP) and mGluR6-GFP to amplify the weak native GFP signal and reveal axonal processes, Rabbit anti-dsRed (1:1000 dilution) (632496, Clontech), Rabbit anti-Calbindin D-28k (1:2000 dilution) (CB38a, Swant), Mouse anti-Calretinin (1:2000 dilution) (MAB 1568, Millipore).

Secondary antibodies raised against the relevant species were obtained from Jackson ImmunoResearch or Invitrogen.

Additional Information for Retinal In Vivo Experiments and Analyses

In Vivo Retinal Electroporation

Postnatal day 0 to 5 (P0-P5) mouse pups were electroporated in vivo as described previously (Matsuda and Cepko, 2004), except that a Femtojet Express pressure injector (Eppendorf;920010521) delivered the DNA via a custom made glass needle (Origio, C060609). DNA solutions (1-1.5 μg/μl) were injected through the sclera and into the subretinal space of the mouse retina. pCAGEN served as an empty vector substitute for excluded plasmids. For demonstration of GFP as a spatial regulator of gene expression in CD1 retinas, GFP, DBDG, and ADG plasmids were electroporated at following DNA mass ratios: CAG-GFP:DBDG:ADG – 3:2.5:1,

Rho-GFP:DBDG:ADG - 3:2.5:1.25, mGluR6-GFP:DBDG:ADG - 4:2.5:1.25. For Tg(GUS8.4GFP) experiment, DBDG:ADG DNA mass ratio was kept at 2.5:1 DNA mass ratio.

Quantification of Electroporated Retinas

20 μm thick retinal cryosections were immunostained for chicken anti-βgal and imaged via confocal microscopy. Regions of dense electroporation were selected for quantification. Quantifications were done on ImageJ using the Cell Counter plugin. The Imaris program aided analyses. To begin, cells expressing a given marker were randomly chosen throughout the confocal stack. Selected cells were then analyzed for co-expression with additional genes, represented by co-localization of fluorescent signals emitted by the chosen channels. A gene was judged, by eye, to be expressed in a cell if its corresponding fluorescent signal was enriched within the selected cell body relative to the surrounding. Individual slices of a confocal stack were analyzed for this purpose. We detected very little n-βgal in the INL even when GFP and tdT were strongly expressed in the layer. This is likely due to a combination of poor expression and/or poor antibody detection. Thus, we limited our analyses n-βgal+ marked cells in the ONL. To sample the mean of each retina, at least 50 and 20 cells were quantified for each marker type in the ONL and INL, respectively. The mean of each parameter was averaged across retinas and plotted as mean ± SEM.

The coelectroporation efficiency between two plasmids was estimated by measuring the percentage of CAG-GFP+ cells per n-βgal+ (from CAG-nlacZ expression) cell in the ONL. This was found to be 91 ± 2% (mean ± SD) (n = 3 retinas). Thus, the probability that 4 plasmids will be electroporated into a cell is approximately $0.91^3 = 0.75$. We divided the non-adjusted efficiency of tdT activation (quantified in bottom of Table S6) by this probability (in percentage). This produces the adjusted efficiency of the T-DDOG system.

In retinas strongly electroporated with DNA including mGluR6-GFP and Gal4-GBP1$^{p65\text{-}GBP6}$, some leakage of mGluR6 promoter was detected in the ONL, giving weak GFP fluorescence. Leakage from mGluR6 promoter was likely, since a mGluR6-ERT2-CreERT2 construct also activated floxed reporter expression in the ONL under similar electroporation conditions. However, leaky GFP was not detected when mGluR6-GFP was electroporated without Gal4-GBP1$^{p65\text{-}GBP6}$ components, even after GFP antibody staining. The appearance of leaky ONL GFP may thus be due to a combination of the fluorescence enhancing ability of GBP1 (Kirchhofer et al., 2010) and/or slight GFP stabilization by T-DDOG components. By immunostaining GFP with a polyclonal GFP antibody, and imaging the antibody signal at wavelengths distinct from the GFP emission range, we detected stabilized GFP expression only in the presence of Gal4-GBP1$^{p65\text{-}GBP6}$ but not Gal4-GBP2$^{p65\text{-}GBP7}$. This suggests that Gal4-GBP1$^{p65\text{-}GBP6}$, but not Gal4-GBP2$^{p65\text{-}GBP7}$, can have some stabilizing effect on GFP. Nevertheless, the leaky GFP fluorescence was barely above background and induced little to no UAS-tdT expression. Due to the difficulty distinguishing the weak GFP signals from background noise, it was difficult to quantify the distribution of GFP/tdT expressing cells in the ONL cell count. We thus chose to represent the enrichment of tdT/GFP expression in the INL relative to the ONL by comparing the fluorescent intensity of the fluorescent proteins in the ONL and INL. Owing to the low background in tdT fluorescent channels, we were able to identify weak tdT+ cells in the ONL with better confidence, so we randomly chose tdT+ cells from both the ONL and INL, quantified fluorescent intensity within cell bodies in single confocal stacks, and measured the corresponding GFP signal within the selected cell bodies. Arbitrary units were obtained as in OTX2 quantification analysis, with the background control being obtained directly next to the chosen cell body.

ON/OFF Bipolar Cell Identification

ON/OFF bipolar cells were distinguished by their axonal stratification in the inner plexiform layer (IPL), following criteria described previously (Ghosh et al., 2004). 20 μm thick retinal sections were stained for rabbit anti-GFP and goat-anti-ChAT, followed by confocal imaging. Confocal stacks of retinal sections were used for analysis. Electroporated GFP+ or GFP+/tdT+ bipolar cells were traced along their axonal processes to the IPL, where the identification of ON/OFF cell was based on the processes' location relative to cholinergic bands labeled by anti-ChAT. ON bipolar processes were those that fell below the midline between the two cholinergic bands (Ghosh et al., 2004).

Otx2 Removal Experiment

Ex Vivo Retinal Electroporation

P0 *Otx2*$^{fl/fl}$ (Tian et al., 2002) were electroporated ex vivo with pCAG-GBP1-10 gly-Gal4DBD (100ng/μl), pCAG-p65AD-GBP6 (50ng/μl), pUAS-Cre (40ng/μl), pCAG-nlacZ (100ng/μl). pCAG-GFP (100ng/μl) or pCAG-dsRed (100ng/μl) was used depending on the experimental conditions. Ex vivo electroporation was carried out as described (Emerson and Cepko, 2011) except that retinas were cultured ex vivo for 8 days before harvesting.

Otx2 Expression Quantification

Electroporated cells labeled by either GFP, dsRed, and/or n-βgal were quantified for OTX2 antibody staining. Using ImageJ, a mask for the chosen cell body was created on the GFP, dsRed, or n-βgal channel and transferred to the OTX2 immunofluorescence channel. Measurements of the mean pixel intensity on single confocal slices were taken for the chosen cell bodies. Each measurement was paired with a background fluorescence intensity measurement from the same slice, chosen at a location in the slice where no cells were present. Although the best background control is one taken right next to the cell, we were unable to do this because OTX2 staining is quite widespread and we could not tell whether faint staining signals on the retina slice were real or not. Our choice of background thus gives the impression that OTX2 is still present at a low level above background, but in fact the real baseline level is above 0. Each measurement was then subtracted from the background to obtain the net fluorescence. As a control,

non-electroporated cells, as determined by a lack of GFP, dsRed or n-βgal expression, were also measured for OTX2 immunostaining level. 20 measurement pairs were made for each set of condition per retina. Measurements were normalized by dividing the intensity values from each electroporated or non-electroporated cell by the average intensity of the non-electroporated cells from the same confocal stack. Values were then multiplied by 100 to obtain the arbitrary unit (a.u.). The average OTX2 immunostaining fluorescence intensity was plotted as a histogram binned at 10 a.u. per bar. 2-tailed Student t test assuming unequal variance was used to test for significance, judged to be $p < 0.05$.

PAX6 ONL Quantification

Confocal images were used to quantify number of PAX6 immunopositive cells in the ONL of electroporated and non-electroporated retinal patches. Electroporated retinal patches were identified by the presence of GFP or dsRed expression, and non-electroporated retinal patches were identified by the absence of GFP or dsRed expression. Quantification were performed over 12 slices of confocal images at an interval of 1 μm/slice. A total of 10 confocal stacks taken from 5 retinas were analyzed per condition. For each retina, two confocal stacks were taken from two different regions separated by at least 200 μm. The identification of ONL was aided by DAPI staining. Counting of PAX6-positive cells in the ONL was aided by the cell counter plugin in ImageJ. Results are presented as a boxplot which depicts the median as the central bar in the box, and the maximum and minimum values as the whiskers. To test for statistical significance, 2-tailed Student t test was performed, assuming unequal variance. Statistical significance occurs when $p < 0.05$.

Brain Recordings and Structural Analyses

In Utero Electroporation

Embryonic day (E)15.5 timed-pregnant female C57BL/6 mice were deeply anesthetized with 2% isoflurane. Uterine horns were exposed and periodically rinsed with warm PBS. Plasmids (1-1.5 μg/μl) were injected into the lateral ventricle of one cerebral hemisphere. Five voltage pulses (50 V, 50 ms duration, 1 Hz) were delivered with round plate electrodes (CUY21 electroporator, NEPA GENE, Japan). Injected embryos were placed back and allowed to mature to delivery.

Slice Preparation

Acute brain slices were obtained from electroporated mice (P8-14) using standard procedures. Mice were deeply anesthetised with isoflurane. The brain was quickly removed and placed into chilled choline-based cutting solution containing (in mM) NaHCO3 (25), NaH2PO4 (1.25), KCl (2.5), MgCl2 (7), glucose (25), CaCl2 (1), choline chloride (110), ascorbic acid (11.6), and pyruvic acid (3.1). The brain was cut coronally at 300μm thickness using a Leica VT1000S vibratome (Leica Instruments, Nussloch, Germany). Slices were placed in a glass chamber filled with artificial cerebrospinal fluid (ACSF) containing (in mM) NaCl (127), KCl (2.5), NaHCO3 (25), NaH2PO4 (1.25), CaCl2 (2), MgCl2 (1) and glucose (25). Cutting and ACSF solutions were saturated with 95% O_2 and 5% CO_2. Slices were incubated at 34°C for 30 min before the experiments.

Electrophysiology

Individual slices were transferred to a recording chamber mounted on an upright microscope (Olympus BX51WI) and continuously superfused (2–3 ml/min) with artificial cerebrospinal fluid at room temperature. Cells were visualized through a 60X water-immersion objective with either infrared differential interference contrast optics or epifluorescence to identify GFP /tdT$^+$, GFP$^+$/tdT , and GFP$^+$/tdT$^+$ neurons. Whole-cell current clamp recordings were obtained from cortical layer 2/3 pyramidal neurons in the region of dense electroporation at P8-12. Patch pipettes (2.5–3.5 MΩ) pulled from borosilicate glass (G150F-3, Warner Instruments) were filled a K+ based internal solution, composed of 135 mM KMeSO4, 5 mM KCL, 5 mM HEPES, 4 mM MgATP, 0.3 mM NaGTP, 10 mM Na2-CrePO4 (pH 7.4). Hyperpolarizing and depolarizing 1,000 ms-long step currents were injected to measure the input resistance, membrane time constant, and action potential responses of each cell.

Two-Photon Imaging

Two-photon laser-scanning microscopy (2PLSM) was performed using a custom microscope as previously described (Carter and Sabatini, 2004). Cell morphology was visualized at P12-14 using Alexa Fluor 594 (10–20 μM) excited with 810 nm light and introduced into the cell through whole-cell recording. A 2PLSM 3-dimensional image stack through each neuron was collected, followed by stacks through 2-3° dendritic segments at a 10-fold higher magnification. For dendritic spine density and spine shape analyses, cells were analyzed used custom-written routines in MATLAB. The size of each spine was measured in the optical section of maximal brightness and dendritic length was estimated from Z projections.

Data Acquisition and Analysis

Membrane voltage deflections were amplified and low-pass filtered at 3 kHz using a Multiclamp 700B amplifier (Molecular Devices), digitized at 10 kHz and acquired using National Instruments acquisition boards and a custom version of ScanImage written in MATLAB (Mathworks) (Pologruto et al., 2003). Electrophysiology and imaging data were analyzed offline using Igor Pro (Wavemetrics) and ImageJ (NIH). In figures, current-clamp traces represent a single acquisition. Data (reported as mean ± SEM) were compared statistically using one-way ANOVA. $p < 0.05$ were considered statistically significant.

Fixed-Tissue Preparation and Imaging

Mice were transcardially perfused with 4% paraformaldehyde and brains were post-fixed for 2 days. Brains were sectioned coronally at 60μm using a Vibratome. No immunoenhancement was used to increase the signal of native fluorophores. Images were acquired with an Olympus VS110 microscope and Olympus FV1000 laser scanning confocal microscope (Harvard Neurobiology Imaging Facility).

Optogenetic Recordings of Mouse Retina

Electroporation

Electroporation was done as stated in the in *vivo* electroporation section. In addition to the T-DDOG components and 10xUAS-ChR2-mCherry, we introduced 5xUAS-tdT in order to boost the fluorescence signal in the retina. This was done to aid the identification of electroporated patches for recordings. CAG-nlacZ was also included as an electroporation marker.

Preparation of Retinas

Retina isolation was done under red illumination (peak wavelength at 580 nm) in Ringer's medium (110 mM NaCl, 2.5 mM KCl, 1 mM $CaCl_2$, 1.6 mM $MgCl_2$, 10 mM D-glucose, 22 mM $NaHCO_3$, bubbled with 5% CO_2/95% O_2, pH 7.4). The retinas were then mounted ganglion cell-side up on filter paper (Millipore, MA) that had a 4-5 mm diameter rectangular aperture in the center, and superfused in Ringer's medium at 35–36°C in the microscope chamber for the duration of the experiment.

Targeting Cells in Wholemount

We visualized red and green fluorescent bipolar cells by using a custom made 2-photon microscope equipped with a Mai Tai HP two-photon laser (Spectra Physics) tuned to 920 nm. Infrared illumination allowed the visualization of the retina with a CCD camera (Diagnostic Instruments) during two-photon scanning. GCL cells, which had fluorescent bipolar cells in their dendritic field were targeted by patch electrodes. A small area above the GCL cells was cleaned from Müller cell end feet using infrared imaging with a patch pipette filled with mouse Ringer's solution. Then the patch electrode was moved into close proximity of the targeted cell and loose cell-attached mode was established. (http://www.youtube.com/watch?v=Epfpnh1jxaU).

Electrophysiology and Pharmacology

Electrophysiological spike recordings were made using an Axon Multiclamp 700B amplifier (Molecular Devices) and borosilicate glass electrodes (Sutter Instrument). Signals were digitized at 10 kHz (National Instruments) and acquired using custom software written in LabVIEW (National Instruments). Data were analyzed offline using MATLAB (MathWorks). The spiking responses were recorded using the patch clamp technique in loose cell-attached mode with electrodes pulled to between three and five MΩ resistance and filled with Ringer's medium. In order to visualize GCL cells in some experiments after spike recordings, neurobiotin (Vector Laboratories) and Alexa 488 (Molecular Probes) were delivered using patch pipette, pulled to between five and eight MΩ resistance and filled with 112.5 mM $CsCH_3SO_3$, 1 mM $MgSO_4$, 7.8×10^{-3} mM $CaCl_2$, 0.5 mM BAPTA, 10 mM HEPES, 4 mM ATP-Na_2, 0.5 mM GTP-Na_3, 5 mM lidocaine N-ethyl bromide (QX314-Br), 7.5 mM neurobiotin chloride, 13 mM Alexa 488. The pH was adjusted to 7.2 with CsOH. In pharmacological experiments, 20 μM APB (L-(+)-2-amino-4-phosphonobutyric acid, blocking metabotropic glutamate receptors) (Calbiochem) was bath-applied.

Light Stimulation

Photoreceptors and were stimulated with light generated by a digital light processor (DLP) projector at 75 Hz (V-339 PLUS Vision Corp.). The same DLP projector provided the infrared light for patch clamp recordings. The maximum power produced by the projector was 229 ± 35 mW/cm^2 (mean ± SEM). We measured light projected through the objective lens at the focal plane on the stage. Light intensity was measured with a photodiode power meter (in Watts per cm^2; Model S130VC; Thorlabs), and the spectrum was measured with a spectrometer (Model USB4000-UV-VIS; Ocean Optics). We have expressed light intensity in number of photoisomerizations per rod per second (R*/s). The light path was computer-controlled with a shutter (SC10, Thorlabs). Light intensity was modulated by neutral density filters, which were built into two filter wheels (FW102, Thorlabs).

For photoreceptor stimulation, the light was focused on the outer segments of photoreceptors in the wholemount retina. For photoreceptor stimulation the light intensity was 1.3×10^3 R*/s. The contrast for photoreceptor stimulation was 2000. For ChR2 stimulation the light was focused on the bipolar cell layer. ChR2/H134R was activated with a 2 s light flash at ~10^8 R*/s generated by a 120 W mercury epifluorescent lamp (X-Cite 120 PC, Lumen Dynamics). The stimulus was generated using software written in LabVIEW (National Instruments) and Python.

Postrecording Immunohistochemistry

For whole-mount immunohistochemistry, fixed retinas were freeze/thawed at least 3 times in 30% sucrose/PBS solution. Retinas were then blocked in 10% normal donkey serum, 1% BSA and 0.35% PBT. Antibody stainings were done in 3% normal donkey serum, 1% BSA, and 0.4% PBT. Primary antibody incubations were done at room temperature for 3-4 days. Retinas were washed with PBS, treated with secondary antibody and washed again with PBS. Retinas were then stained for neurobiotin using streptavidin-Alexa 649 probes in 0.5% PBT overnight at room temperature. Washed retinas were flat-mounted and imaged with confocal microscopy. Antibody stainings were done for anti-ChAT to analyze stratification and anti-dsRed to boost mCherry/tdT signals.

Screens

T-DDOG Screen

A total of 6 GBPs (GBP1, 2, 4, 5, 6, and 7) were used for all T-DDOG screens. Each GBP was fused to a DBD or an AD at either its N- or C-terminal end. Fusions were made with either no linker between the two protein modules, or with a 10 amino acid glycine linker. Most chimeric constructs carried a nuclear localization signal (NLS) at the N-terminal end. All chimeric constructs were placed under the control of the ubiquitous CAG promoter. The resulting CAG-chimera plasmids were combined in many possible pairwise combinations and transfected along with pCAG-GFP, UAS-luc2 and pRL-TK into 293T cells. Combinations that gave the strongest reporter induction after 24 hr were selected for further characterization.

rTetR-based T-DDOGs were screened similarly as above, but transfection mixtures for each DBDG:ADG combination was split into two culture wells, one with 1 μg/ml doxycycline and the other with no doxycycline. Functional combinations were judged to be those that gave strong reporter induction in the presence of doxycycline, but not in the absence of doxycycline.

Rod Mispositioning Screen

In our initial attempts to apply Gal4-GBPX$^{VP16\text{-}GBPY}$ in the retina, we observed that T-DDOGs with VP16AD induced mispositioning of rod photoreceptors in the upper (sclera) half of the ONL. This phenotype was seen in retinas electroporated with VP16AD-GBPX, but not with Gal4DBD-GBPX plasmids. The phenotype also was independent of UAS-transgene activation or Gal4-GBPX$^{VP16\text{-}GBPY}$ interaction with GFP, since we observed the same phenotype in electroporated cells not expressing GFP and UAS-tdT, but expressing Gal4-GBPX$^{VP16\text{-}GBPY}$. Thus, high levels of VP16AD likely played a major role in the mispositioning of rod photoreceptors. The mechanism behind this is likely due to squelching effect of potent activation domains (Gill and Ptashne, 1988).

To overcome this problem, we screened for ADs that did not give the mispositioning phenotype. The most useful ADs would be those that confer high transcriptional activity with minimal side effects on cellular phenotype. To determine the best AD for use in the retina, selected T-DDOGs involving p65 or VPmin ADs were electroporated into P0 CD1 retina along with CAG-GFP, UAS-tdT and CAG-nlacZ. We chose from T-DDOGs that gave strong GFP-dependent induction of UAS-luc2 in 293T cells. Electroporation mixtures including VP16AD served as the positive control for the rod phenotype, while those missing DBDG and ADG plasmids served as the negative control. DNA mass amount was kept constant for all conditions, with GFP:DBDG:ADG plasmid ratio at 2:2:1.

Of the 13 T-DDOGs tested, those utilizing p65 as the AD consistently gave the most normal positioning of rod photoreceptors. T-DDOGs utilizing VPminx2 also had reduced mispositioning phenotype, but their activities were also relatively weaker than that of p65. T-DDOGs utilizing VPminx3 or x4 strongly induced UAS-tdT activity, but they also caused obvious photoreceptor mispositioning. Given these results, we chose to work with p65AD in our in vivo/ex vivo experiments. It should be noted that whether a particular AD will cause abnormality in the host cell depends largely on the context, such as cell type identity, method of T-DDOG delivery, expression level of T-DDOG components, etc. This is well documented in other gene regulatory systems using transcription factors (Asakawa et al., 2008; Kramer and Staveley, 2003; Morimoto and Kopan, 2009; Pfeiffer et al., 2010; Potter et al., 2010). For example, we found that electroporation of high amounts of p65AD plasmids in similar DNA mixtures eventually led to rod mispositioning. Thus, we stress the importance of always performing proper controls to rule out potential side effects caused by overexpression of AD.

SUPPLEMENTAL REFERENCES

Asakawa, K., Suster, M.L., Mizusawa, K., Nagayoshi, S., Kotani, T., Urasaki, A., Kishimoto, Y., Hibi, M., and Kawakami, K. (2008). Genetic dissection of neural circuits by Tol2 transposon-mediated Gal4 gene and enhancer trapping in zebrafish. Proc. Natl. Acad. Sci. USA *105*, 1255–1260.

Ballard, D.W., Dixon, E.P., Peffer, N.J., Bogerd, H., Doerre, S., Stein, B., and Greene, W.C. (1992). The 65-kDa subunit of human NF-kappa B functions as a potent transcriptional activator and a target for v-Rel-mediated repression. Proc. Natl. Acad. Sci. USA *89*, 1875–1879.

Carter, A.G., and Sabatini, B.L. (2004). State-dependent calcium signaling in dendritic spines of striatal medium spiny neurons. Neuron *44*, 483–493.

Emerson, M.M., and Cepko, C.L. (2011). Identification of a retina-specific Otx2 enhancer element active in immature developing photoreceptors. Dev. Biol. *360*, 241–255.

Huang, L., Shanker, Y.G., Dubauskaite, J., Zheng, J.Z., Yan, W., Rosenzweig, S., Spielman, A.I., Max, M., and Margolskee, R.F. (1999). Ggamma13 colocalizes with gustducin in taste receptor cells and mediates IP3 responses to bitter denatonium. Nat. Neurosci. *2*, 1055–1062.

Kramer, J.M., and Staveley, B.E. (2003). GAL4 causes developmental defects and apoptosis when expressed in the developing eye of *Drosophila melanogaster*. Genet. Mol. Res. *2*, 43–47.

Kurosu, T., and Peterlin, B.M. (2004). VP16 and ubiquitin; binding of P-TEFb via its activation domain and ubiquitin facilitates elongation of transcription of target genes. Curr. Biol. *14*, 1112–1116.

Lewis, B.P., Shih, I.H., Jones-Rhoades, M.W., Bartel, D.P., and Burge, C.B. (2003). Prediction of mammalian microRNA targets. Cell *115*, 787–798.

Morimoto, M., and Kopan, R. (2009). rtTA toxicity limits the usefulness of the SP-C-rtTA transgenic mouse. Dev. Biol. *325*, 171–178.

Pfeiffer, B.D., Ngo, T.T., Hibbard, K.L., Murphy, C., Jenett, A., Truman, J.W., and Rubin, G.M. (2010). Refinement of tools for targeted gene expression in Drosophila. Genetics *186*, 735–755.

Pologruto, T.A., Sabatini, B.L., and Svoboda, K. (2003). ScanImage: flexible software for operating laser scanning microscopes. Biomed. Eng. Online *2*, 13.

Potter, C.J., Tasic, B., Russler, E.V., Liang, L., and Luo, L. (2010). The Q system: a repressible binary system for transgene expression, lineage tracing, and mosaic analysis. Cell *141*, 536–548.

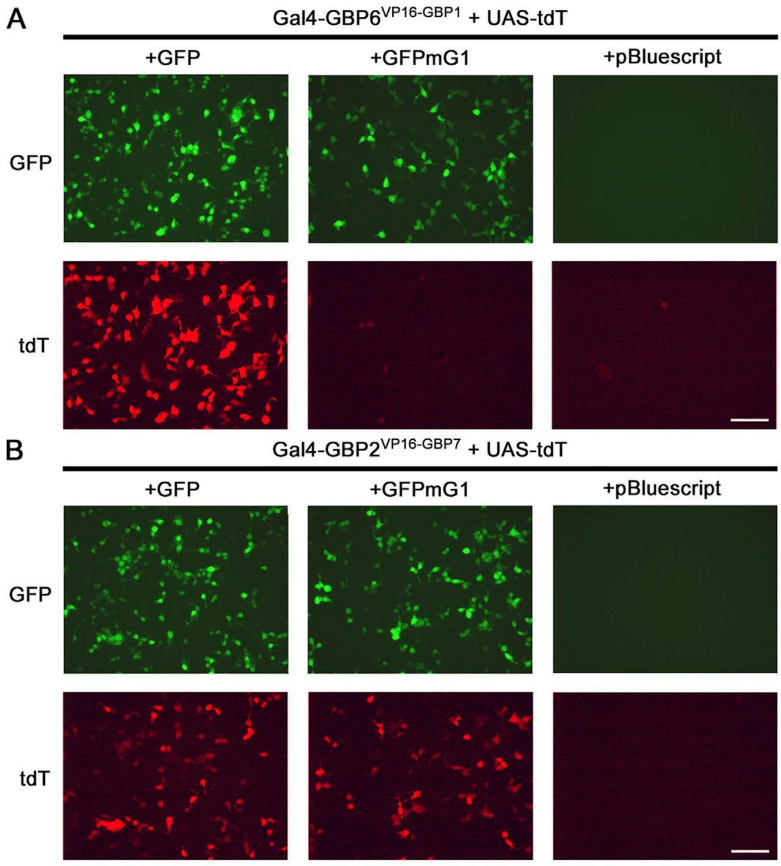


Figure S1. GFP-Dependent Activation of UAS-tdT In Vitro, Related to Figure 2.2

(A and B) Representative images of Gal4-GBP6$^{VP16\text{-}GBP1}$ (A) and Gal4-GBP2$^{VP16\text{-}GBP7}$ (B) activation of an UAS-tdT reporter in 293T cells. Cells were transfected with all indicated constructs and imaged 16 hr post-transfection. GFP, GFPmG1, DBDG and ADG constructs were delivered in separate plasmids under the CAG-promoter. Results are representative of 4 independent experiments and quantifications are tabulated in Table S2. Scale bar, 100 μm.

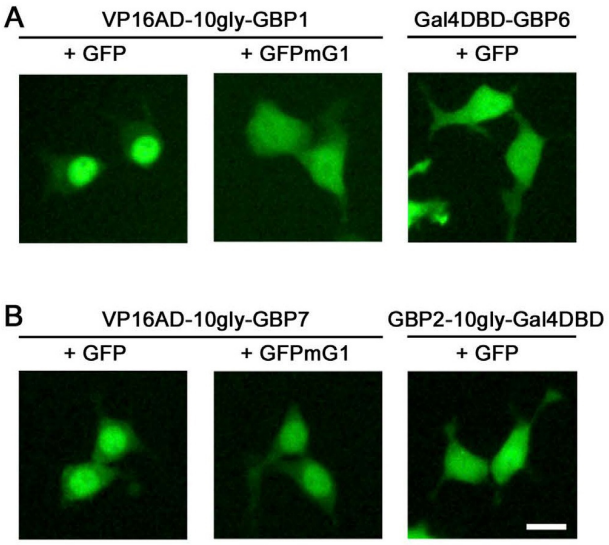


Figure S2. VP16AD-GBP Fusions Can Localize GFP to the Nucleus, Related to Figure 2.2

Representative images of GFP localization in 293T cells transfected with pCAG-GFP or pCAG-GFPmG1 along with CAG-driven DBDG or ADG plasmids at a 1:2 molar plasmid ratio. Fluorescent micrographs were taken at 16 hr post-transfection in live cells. Scale bar, 15μm.

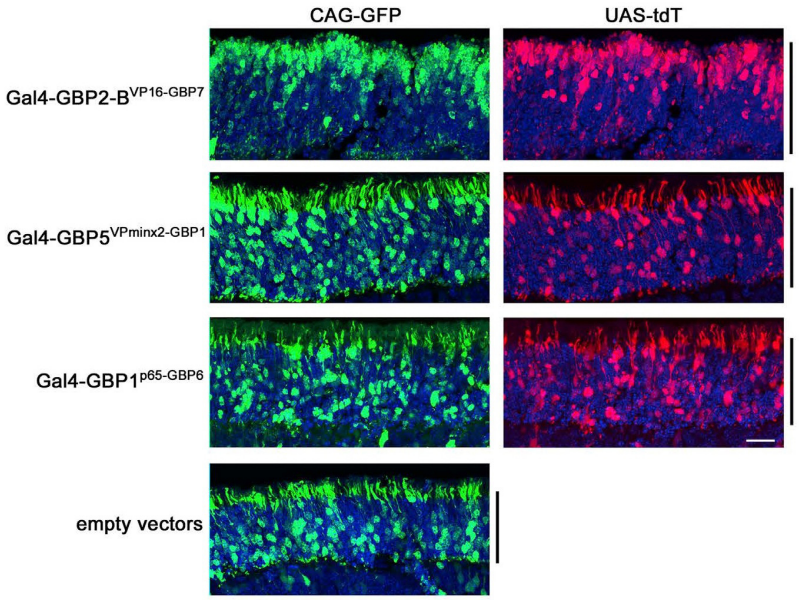


Figure S3. T-DDOGS Composed of VPminx2AD and p65AD Give Normal Positioning of Rod Photoreceptors in the Retina, Related to Figure 2.4
Representative images showing the effects of selected T-DDOGs on rod photoreceptor positioning. CD1 mouse retinas were electroporated at P0 in vivo with plasmids encoding CAG-GFP (green), Gal4-GBPX$^{AD-GBPY}$ components, UAS-tdT (red) and CAG-nlacZ. The electroporated GFP:DBDG:ADG plasmid mass ratio is 2:2:1. Retinas were harvested at P14. Confocal stacks of retina sections spanning the ONL (black bar) are shown. Each set of panels was adjusted individually to similar levels to reveal the cell bodies. DAPI is blue in all panels. Scale bar, 20 μm. Non-linear adjustment using the gamma function was applied to show weak GFP or tdT-expressing cells to minimize their signals from being obscured by those of stronger-expressing equivalents. The scleral portion of the ONL is the upper half of each image.

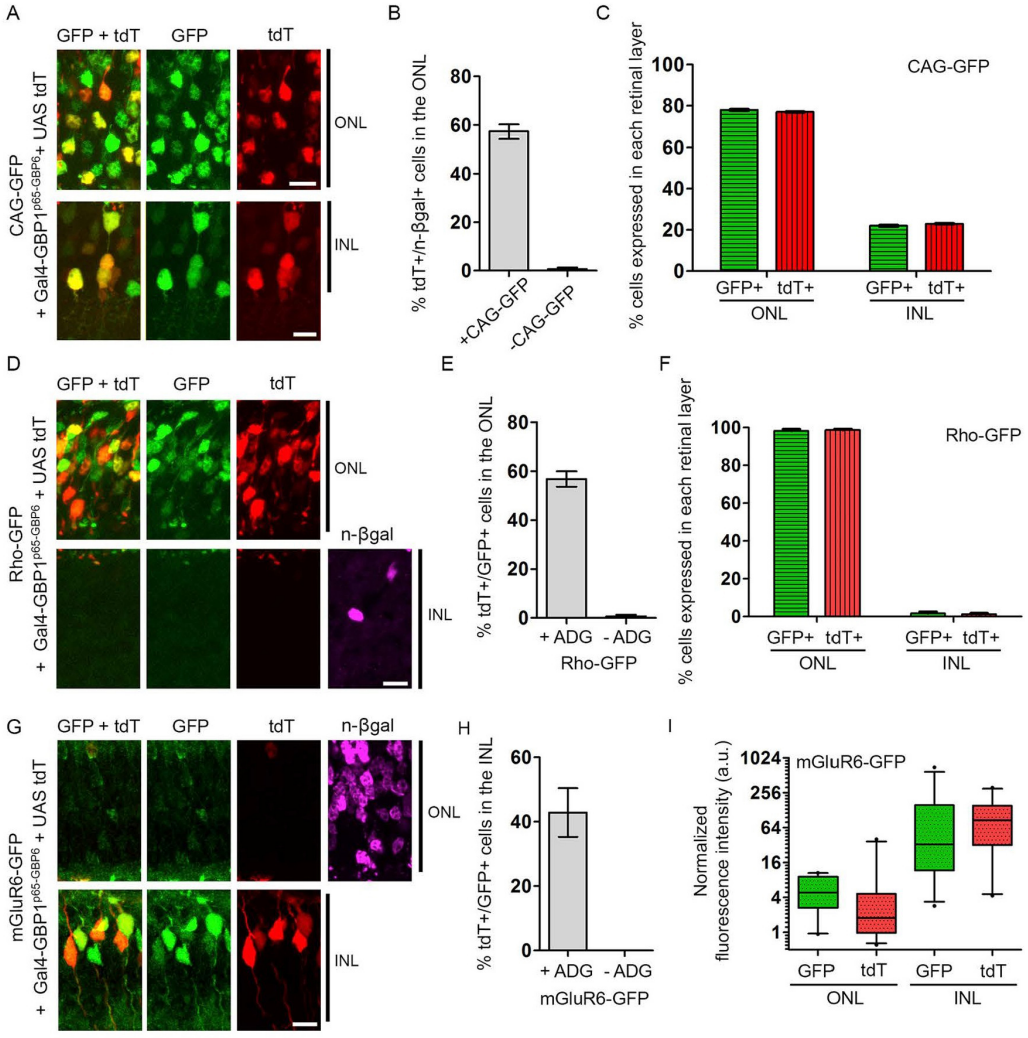


Figure S4. Quantification of Gal4-GBP1$^{p65\text{-}GBP6}$ Activities in the Retina, Related to Figure 2.4

Refer to Figure 4 and Experimental Procedures for experimental setup.

(A, D, and G) Colocalization of different promoter-driven GFP (green) with tdT (red) in retinas expressing Gal4-GBP1$^{p65\text{-}GBP6}$.

(B, E, and H) Efficiency and specificity of Gal4-GBP1$^{p65\text{-}GBP6}$ in vivo. Removal of CAG-GFP (B) or ADG (E, H) abolished reporter expression. In (B), the ordinate shows the percentage of tdT-expressing cells among the electroporated cells in the ONL, as assessed by expression of the electroporation marker, n-βgal. n-βgal detection and/or expression were poor in the INL, preventing analysis of Gal4-GBP1$^{p65\text{-}GBP6}$ efficiency in the INL from the sample in (B).

(C and F) Distribution of GFP and tdT expressing cells in electroporated retinas. Weak, leaky mGluR6-GFP expression in the ONL made it difficult to perform this analysis by cell-counting. (I) Quantification of fluorescent intensity of cells expressing GFP and tdT in (G). Boxplots indicate median, interquartile range, and 5%–95% range. *y* axis is on log2 scale. n = 25 each condition. For (B), (C), (E), (F), (H), plots are mean and SD. Scale bar, 10 μm. Note the tdT+/GFP+ data points for Rho-GFP, +ADG (E) and mGluR6-GFP, +ADG (H) were also presented in Figure S5 for a different argument.

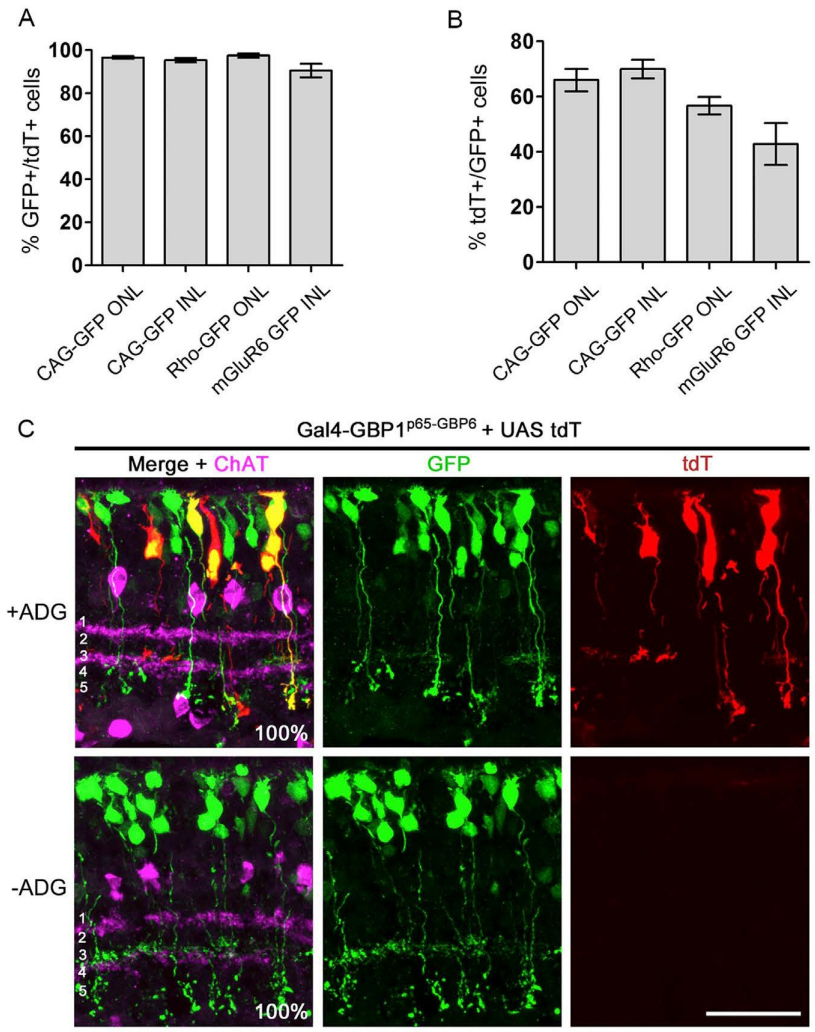


Figure S5. Additional Analysis of Electroporated Retinas from Figure S4, Related to Figure 2.4

Refer to Figure 4 and Methods for experimental setup.

(A) Tight-coupling of tdT+ cells with GFP expression. Among cells that expressed tdT, the percentage of cells expressing GFP is shown.

(B) Efficiency of tdT activation in GFP-expressing cells. Percentage of tdT+ cells among GFP+ cells. Note the tdT+/GFP+ data points for Rho-GFP and mGluR6-GFP were also presented in Figure S4 for a different argument. Plots show mean ± SD. See Table S3 for summary.

(C) mGluR6-GFP bipolar cells project to the correct IPL sublaminae in the presence of Gal4-GBP1$^{p65\text{-}GBP6}$. CD1 mouse retinas were electroporated at P0 with plasmids encoding mGluR6-GFP (green), Gal4-GBP1$^{p65\text{-}GBP6}$, UAS-tdT (red) and CAG-nlacZ. Retinas stained for anti-GFP (green) and anti-ChAT (magenta). The IPL is divided into 5 sublaminae (white font in merge panels), with the ON bipolar cells stratifying below the midline dividing sublaminae 2 and 3. Anti-ChAT labels the boundary between sublaminae 1 and 2, and also the boundary between 3 and 4. % GFP+ and GFP+/tdT+ cells stratifying below the midline of 2 and 3 are shown in merge panels. n = 28 for +ADG condition and n = 31 for –ADG condition. Cells were sampled across two retinas per condition. Scale bar, 20 μm.

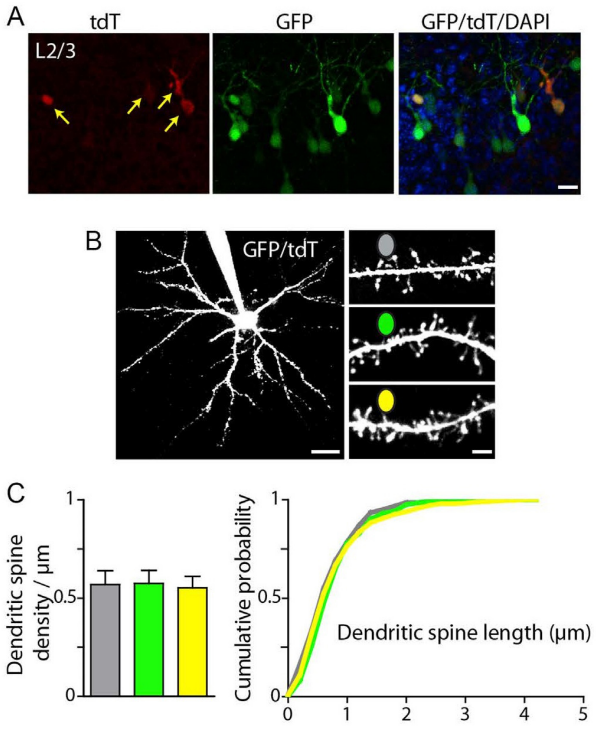


Figure S6. Structural Analysis of Electroporated Cortical Pyramidal Neurons, Related to Figure 2.5

Primary somatosensory cortex was electroporated at E15.5 with plasmids encoding GFP, ADG, DBDG, and UAS-tdT. T-DDOG is Gal4-GBP1$^{p65\text{-}GBP6}$.

(A) Confocal image of native GFP and tdT fluorescence in layers 2/3 of mouse primary somatosensory cortex. Red, tdT; green, GFP; overlay shows both channels with DAPI. A subset of GFP+ cells are tdT+, reflecting co-electroporation of 4 constructs. Arrows point to GFP+/tdT+ cells. Scale bar, 20 µm.

(B) Left, 2-photon Z-stack through a GFP+/tdT+ layer 2/3 pyramidal neuron filled with Alexa-594 through a patch pipette. Right, dendrites from GFP-, GFP+/tdT- and GFP+/tdT+ neurons at P13. Scale bars, 30 and 3 µm.

(C) Summary data of dendritic spine density (left) and length (right) for electroporated pyramidal neurons at P12-14. Dendritic spine density and length did not differ among GFP-, GFP+/tdT- and GFP+/tdT+ neurons. (Dendritic spine density, n = 7 neurons and 15-24 dendrites per condition; spine length, ~300 dendritic spines per condition). No GFP-/tdT+ neurons were observed in acute slices during recording or 2-photon imaging (n = 7 mice). Column plots are mean ± SEM.

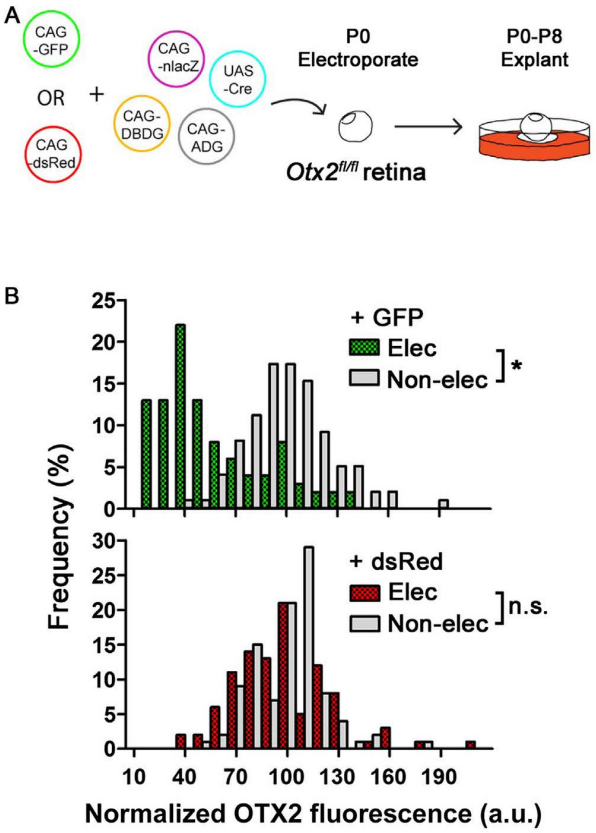


Figure S7. Excision of *Otx2*$^{fl/fl}$ Led to Decrease in OTX2 Protein Levels, Related to Figure 2.5

(A) Schematic of experiment. T-DDOG is Gal4-GBP1$^{p65\text{-}GBP6}$.

(B) OTX2 protein levels were measured by OTX2 immunofluorescence in single electroporated (Elec), GFP+ or dsRed+/n-βgal + cells in the ONL. Measured values are normalized against that of non-electroporated cells (Non-elec), judged by absence of GFP, dsRed and n-βgal. n = 100 each condition. * indicates $p < 0.0001$. n.s., not significant.

Appendix II

Rejected Cover Art

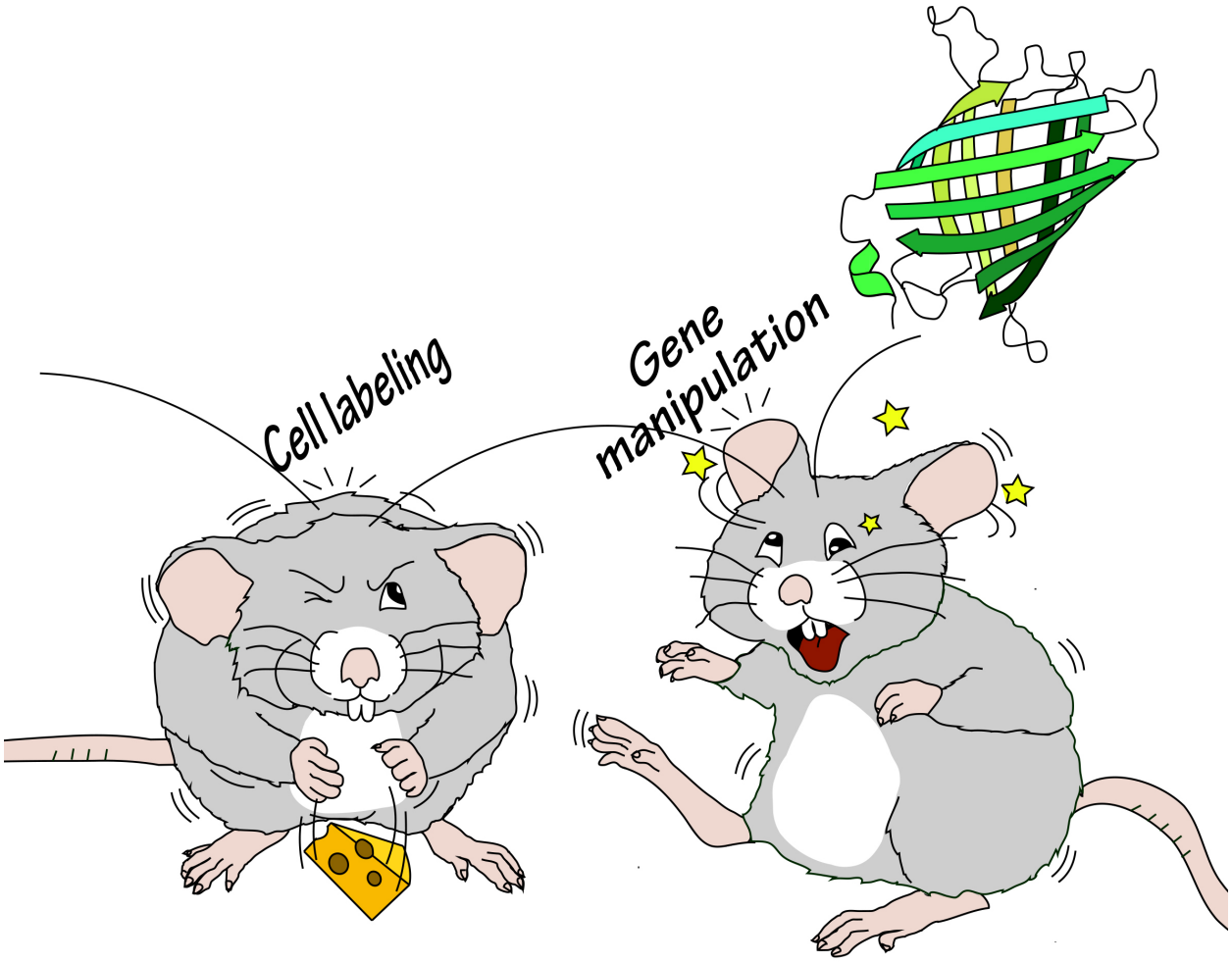


Figure 1. Rejected Cell Cover 1.

Figure 2. Rejected Cell cover 2. See next page.

Figure 2. (Continued)