2015-03-16

A More Accessible Drosophila Genome to Study Fly CNS Development: A Dissertation

Hui-Min Chen
University of Massachusetts Medical School Worcester

A MORE ACCESSIBLE *DROSOPHILA* GENOME TO STUDY FLY CNS DEVELOPMENT

A Dissertation Presented

By

Hui-Min Chen

Submitted to the Faculty of the

University of Massachusetts Graduate School of Biomedical Sciences, Worcester

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

March 16th, 2015

A MORE ACCESSIBLE *DROSOPHILA* GENOME TO STUDY FLY CNS DEVELOPMENT

A Dissertation Presented
By

Hui-Min Chen

The signatures of the Dissertation Defense Committee signify
completion and approval as to style and content of the Dissertation

_____
Tzumin Lee, M.D., Ph.D., Thesis Advisor

_____
David Weaver, Ph.D., Member of Committee

_____
Scot Wolfe, Ph.D., Member of Committee

_____
Michael Brodsky, Ph.D., Member of Committee

_____
Avital Rodal, Ph.D., Member of Committee

The signature of the Chair of the Committee signifies that the written dissertation meets
the requirements of the Dissertation Committee

_____
Marc Freeman, Ph.D., Chair of Committee

The signature of the Dean of the Graduate School of Biomedical Sciences signifies
that the student has met all graduation requirements of the school.

_____
Anthony Carruthers, Ph.D.,
Dean of the Graduate School of Biomedical Sciences

Program in Neuroscience

March 16th, 2015

Dedication

To

My dear parents, for their endless love, trust, and support;

And to my beautiful wife, for her everlasting love, patience, and encouragement that accompany me over my failures, struggles, and little success;

Devoted, to my adorable son, for his genuine love, sincere heart, and spontaneous smiles and laughter;

Also, to the ones, who I love or love me;

To a name that carries the spirits of this project.

Acknowledgements

I would like to thank Tzumin Lee for his acute guidance, forever passion, and constant patience and trust in me. Also, I would like to thank my TRAC members for their continual valuable suggestions, patience, and support though the distance between us. Thanks to Yaling Huang, Barret Pfeiffer, Xiaohao Yao, Ling-Yu Liu, Jon-Michael Knapp, and Fillip Port and for their contributions to the development of Golic+. Thanks to Ching-Po Yang, Josephine Chang, and Haluk Lacin for their contributions to the molecular marker/NB/Lineage study. Thanks to the current and former Lee lab members for their precious comments and suggestions, especially Mark Schroeder, Suewei Lin, Takeshi Awasaki, Hung-Hsiang Yu, Chih-Fei Kao, Lei Shi, Shun-Jen Yang, HaoJiang Luan, Zhiyong Liu, and Ying-Jou Lee. Thanks to the people in the JRC shared resources, especially Todd Laverty, Karen Hibbard, Monti Mercer, Melissa Ramirez, and Phuong Nguyen. Thanks to Dennis M. McKearin, Julian Ng, and Justin Crocker for their valuable reagents. Thanks to Tara Keegan and Colleen Baldelli in UMMS and Crystal Sullivan in JRC for their attentive support and administrative work. And finally, I would like to thank the Department of Neurobiology, UMass Medical School, and Janelia Research Campus, Howard Hughes Medical Institute for the cooperative and inspiring research environments.

Abstract

Understanding the complex mechanisms to assemble a functional brain demands sophisticated experimental designs. *Drosophila melanogaster*, a model organism equipped with powerful genetic tools and evolutionarily conserved developmental programs, is ideal for such mechanistic studies. Valuable insights were learned from research in *Drosophila* ventral nerve cord, such as spatial patterning, temporal coding, and lineage diversification. However, the blueprint of *Drosophila* cerebrum development remains largely unknown.

Neural progenitor cells, called neuroblasts (NBs), serially and stereotypically produce neurons and glia in the *Drosophila* cerebrum. Neuroblasts inherit specific sets of early patterning genes, which likely determine their individual identities when neuroblasts delaminate from neuroectoderm. Unique neuroblasts may hence acquire the abilities to differentially interpret the temporal codes and deposit characteristic progeny lineages. We believe resolving this age-old speculation requires a tracing system that links patterning genes to neuroblasts and corresponding lineages, and further allows specific manipulations.

Using modern transgenic systems, one can immortalize transient NB gene expressions into continual labeling of their offspring. Having a collection of knock-in drivers that capture endogenous gene expression patterns would open the door for tracing specific NBs and their progenies based on the combinatorial expression of various early patterning genes. Anticipating the need for a high-

throughput gene targeting system, we created Golic+ (**g**ene targeting during **o**ogenesis with **l**ethality **i**nhibitor and **C**RISPR/Cas “plus”), which features efficient homologous recombination in cystoblasts and a lethality selection for easy targeting candidate recovery. Using Golic+, we successfully generated T2A-Gal4 knock-ins for 6 representative early patterning genes, including *lab*, *unpg*, *hkb*, *vnd*, *ind*, and *msh*. They faithfully recapitulated the expression patterns of the targeted genes. After preserving initial NB expressions by triggering irreversible genetic labeling, we revealed the lineages founded by the NBs expressing a particular early patterning gene.

Identifying the neuroblasts and lineages that express a particular early patterning gene should elucidate the genetic origin of neuroblast diversity. We believe such an effort will lead to a deeper understanding of brain development and evolution.

Table of Contents

List of Tables

List of Figures

List of Third Party Copyrighted Material

1) Figure 3.1: Embryonic *in situ* images of *lab*, *unpg*, *hkb*, *vnd*, *ind*, *msh* were modified from BDGP insitu database (http://insitu.fruitfly.org).

List of Symbols, Abbreviations, or Nomenclature

AL	antennal lobe
AN	Antennal
AP	anterior/posterior
bamP-CFI	bamP-Cas9-2A-FLP-2A-I-SceI
bHLH	basic helix-loop-helix
BMP	bone morphogenetic protein
BPfife	bigger Pfife
Cas	CRISPR-associated
CB	cystoblast
CDS	coding sequence
CNS	central nervous system
CRISPR	clustered regularly interspaced short palindromic repeats
crRNA	CRISPR RNA
DER	Drosophila EGF receptor
dpn	deadpan
DSB	double-strand break
DV	dorsal/ventral
EGFR	epidermal growth factor receptor
en	*engrailed*
ES	embryonic stem
FF	founder females

FLP	flippase
FRT	FLP recognition target
GMC	ganglion mother cell
Golic+	gene targeting during oogenesis with lethality inhibitor and CRISPR/Cas
gRNA	guide RNA
gsb	*gooseberry*
GSC	germline stem cell
GT	gene targeting
HR	homologous recombination
IC	intercalary
ind	intermediate neuroblasts defective
INP	intermediate neural progenitor
kb	kilobase
KI	knock-in
LA	labial
LH	Lateral horn
LR	labral
MB	mushroom body
MD	mandibular
msh	muscle segment homeobox
MX	maxillary
NB	neuroblast

NHEJ	non-homologous end-joining
ORF	open reading frame
OC	ocular
PAM	protospacer adjacent motif
PAN	posterior Asense-negative
Pfife	p10-facilitated indicators of flip excision
pNR	procephalic neurogenic region
pTL	plasmid, targeting with lethality selection
rCD2i	rat CD2 miRNA
RMCE	recombinase-mediated cassette exchange
RNAi	RNA interference
RVD	repeat variable di-residue
SEG	subesophageal ganglion
SPG	supraesophageal ganglion
TALEN	transcription activator-like effector nuclease
TAP	transient amplifying precursor
tracrRNA	trans-activating crRNA
TS	target site or target sequence
UAS-CFI	UAS-Cas9-2A-FLP-2A-I-SceI
VNC	ventral nerve cord
vnd	ventral nervous system defective
vNR	ventral neurogenic region
WD	wandering

wg	*wingless*
ZFN	zinc-finger nuclease

Preface

1) Chapter II has been published: Chen, H. M., Huang, Y., Pfeiffer, B. D., Yao, X., & Lee, T. (2015). An Enhanced Gene Targeting Toolkit for *Drosophila*: Golic+. Genetics, **199**: 683-694. Hui-Min Chen prepared the manuscript. Tzumin Lee wrote the paper.
2) Part of the work in Chapter II (transgenic *Drosophila* CRISPR/Cas system; using *dU6-2* and *dU6-3* promoter to express gRNA) was also published through collaboration: Port, F., Chen, H. M., Lee, T., and Bullock, S. L. (2014) Optizmized CRISPR/Cas tools for efficient germline and somatic genome engineering in Drosophila. Proc Natl Acad Sci USA, **111**: E2967-2976. Fillip Port and Simon L. Bullock wrote the paper.
3) Tzumin Lee conceived the project of Chapter II. Yaling Huang, Barret Pfeiffer, and Xiaohao Yao provided molecular biology support for a subset of DNA constructs. Hui-Min Chen designed and conducted the experiments, collected and analyzed the data.
4) Chapter III is unpublished. Likely authors: Hui-Min Chen, Ching-Po Yang, Haluk Lacin, Josephine Chang, and Tzumin Lee.
5) Tzumin Lee conceived the project of Chapter III. Hui-Min Chen designed and generated the T2A-Gal4 knock-ins. Josephine Chang made the donor constructs. Ching-Po Yang characterized T2A-Gal4 knock-in expression at the embryonic and larval stages (Figure 3.1). Haluk Lacin characterized T2A-Gal4 knock-in expressions in embryonic ventral neurogenic region (Figure

3.2). Ching-Po Yang performed the immortalization experiments (Figure 3.3, 3.4), and Hui-Min Chen mapped the hit lineages by stochastic clonal labeling.

6) Embryonic in situ images of *lab*, *unpg*, *hkb*, *vnd*, *ind*, *msh* were modified from BDGP insitu database (http://insitu.fruitfly.org) (Figure 3.1).

CHAPTER I

Introduction

The brain is one of the most complex and fascinating organs of a living organism. It consists of numerous types of neurons and glia that support and communicate with each other to direct sophisticated and purposeful behavioral outputs. Currently, it remains largely a mystery how brain cells, each with their own characteristics, wire their fine processes together and further aggregate to form higher orders of functional structures. However, it has become clear that these diverse cells are serially deposited by their parental stem cells from embryogenesis to adulthood. Mapping individual neurons based on their developmental origins offers a straightforward strategy to resolve this complex structure and its development simultaneously.

The spatiotemporal patterne of neurogenesis suggests three mechanisms by which a brain can acquire its full collection of diversified neurons and glia. First, each stem cell can be specified differently when it emerges from a neuroectoderm (Jessell 2000; Bertrand *et al.* 2002). Second, a temporal patterning system can further modify the properties of stem cells and their offspring throughout development (Wonders and Anderson 2006; Molyneaux *et al.* 2007; Franco *et al.* 2012). Third, post-mitotic neurons and glia can be further shaped by the environment to acquire their final identities (Hidalgo *et al.* 2005; da

Silva and Wang 2011). Overall, neurogenesis is plastic and coordinates well with organism growth (Zhao *et al.* 2008; Lin *et al.* 2013). These mechanisms, and likely others, jointly instruct the formation of functional neural and glial networks in a brain.

The brain of fruit fly, *Drosophila melanogaster*, serves as an attractive model nervous system to study these intricate mechanisms in great depth. First, it has well-described compartments and relatively large but still manageable number of neurons (~22,000 neurons per cerebrum) (Rein *et al.* 2002; Jenett *et al.* 2012; Yu *et al.* 2013a). It is composed of distinct neuropil structures, such as the mushroom bodies, central complex, optic lobes, antennal lobes, and other specialized neuropils and fiber tracts, which have no counterparts in the ventral nerve cord (VNC), the posterior truncal part of the fly central nervous system (CNS). A significant amount of research has been done to understand how these structures are serially and stereotypically produced (Lin and Lee 2012). Their origin is of great interest because these neurons function together and generate versatile innate and acquired behaviors (Griffith 2012). Lastly, there exists a more than adequate set of genetic tools to target and manipulate fly neurons and their progenitors throughout development (Lee and Luo 1999; Laı and Lee 2006; Luan *et al.* 2006; Venken and Bellen 2007; Pfeiffer *et al.* 2008; Yu *et al.* 2009; Pfeiffer *et al.* 2010; Venken *et al.* 2011; Pfeiffer *et al.* 2012; Awasaki *et al.* 2014). However, directly studying fly cerebrum development remains a challenging task because of its hidden segmental organization and higher complexity than the

relatively simply structured VNC, where most of the insights into the spatial (Urbach and Technau 2004; Technau *et al.* 2006) and temporal (Jacob *et al.* 2008; Kao and Lee 2010; Truman *et al.* 2010) mechanisms have been derived.

Basic principles for *Drosophila* brain development

Though structurally obscure, insect brains are traditionally subdivided into three parts: tritocerebrum, deutocerebrum, and protocerebrum (Bullock and Horridge 1965), which originate from the intercalary, antennal and ocular/labral head segments, individually (Figure 1.1A) (Rempel 1975; Schmidt-Ott and Technau 1992; Hirth *et al.* 1995; Younossi-Hartenstein *et al.* 1996). Major structural plans of the brain development are laid down during embryogenesis, but it is mostly unknown how CNS is built from stem cells in neuroectoderm to a fully functional neuronal network (Nassif *et al.* 1998; Kurusu *et al.* 2000; Noveen *et al.* 2000).

To decipher these spatiotemporally intertwined developmental codes, it is reasonable to begin with investigating how *Drosophila* neuronal stem cells, termed neuroblasts (NBs), are first specified. The neurogenic regions of the ectoderm are determined in the blastoderm by early dorsal/ventral and anterior/posterior patterning genes (St Johnston and Nüsslein-Volhard 1992; Biehs *et al.* 1996). Then, proneural and neurogenic genes act in neuroectoderm to select and specify NBs (Artavanis-Tsakonas and Simpson 1991; Campos-Ortega 1995). From embryonic stage 8 to 11, NBs of *Drosophila* CNS delaminate from the neuroectoderm (Hartenstein and Campos-Ortega 1984; Doe 1992;

Bossing *et al.* 1996), which is composed of the ventral neurogenic region (vNR) and the procephalic neurogenic region (pNR) (Figure 1.1A) (Hartenstein and Campos-Ortega 1985; Technau and Campos-Ortega 1985). With NBs giving birth to most of the adult neurons over the larval and into the pupal stages (Truman and Bate 1988; Ito and Hotta 1992), vNR and pNR, as a result, give rise to VNC and the brain, respectively.

It is well known that a grid-like coordinate system exists in the vNR to specify repetitive sets of NBs in each segment of VNC (Figure 1.1A) (Doe 1992; Broadus *et al.* 1995). Proven by classic transplantation experiments, NB fates are specified by positional cues before the delamination occurs (Udolph *et al.* 1995). However, such a clean and organized system disappears in the cerebral epithelium. The Technau group had put an epic effort to follow the formation of 106 NBs in the pNR and survey the expression of a collection of “molecular markers” in each of them (Urbach *et al.* 2003; Urbach and Technau 2003b; Urbach and Technau 2003a).

NBs in the brain are created in a stereotypical manner in time and space like their counterparts in VNC (Urbach *et al.* 2003). Distinct expression domains of segment polarity and dorsal-ventral patterning genes reveal the hidden segmental organization of the fly brain (Urbach and Technau 2003b). Further, each brain NBs is born expressing a specific combination of molecular markers, implying a distinct and unique identity for individual NBs (Urbach and Technau

2003a). Such a map provides the initial clues of possible determining factors in NB specification.

Indeed, removing activities of *vnd*, a columnar gene expressed in the ventral domain of both VNC and brain neuroectoderm, leads to improper specification of ventral brain NBs, a phenotype analogous to the ones observed in the VNC (Urbach *et al.* 2006). Yet, the dorsal columnar gene *msh* is only required for dorsal NB specification in the posterior brain (tritocerebrum and deutocerebrum), not the ones in the prominent anterior brain (protocerebrum) (Urbach *et al.* 2006). Also, in total, less than 50% of the brain NBs express any of the three columnar genes (*vnd*, *ind*, and *msh*), which are heavily involved in VNC dorsal/ventral patterning (Urbach and Technau 2003b). Additionally, the fourth columnar gene *DER* does not seem to participate in brain NB development (Dumstrei *et al.* 1998). As for the anterior/posterior axis, roles of the majority of segment polarity genes remain to be explored, while two *Antennapedia-Complex* Hox genes, *proboscipedia* and *labial* (*lab*), have been demonstrated to be expressed in and required for posterior brain patterning (Hirth *et al.* 1998). All in all, we need a more sophisticated strategy to systematically and specifically identify and characterize essential factors for brain NB specification.

Diverse NB lineages in a *Drosophila* brain

Each NB with its uniquely specified identity in succession produces a corresponding discrete neuronal lineage (sequentially born neurons and glia from

a common progenitor) (Figure 1.1B). It undergoes continual asymmetric divisions to self-renew itself and deposits a series of intermediate progenitor cells named ganglion mother cells (GMCs). Most GMCs then divide once to generate two post-mitotic cells, which can either differentiate into neurons, glia, or die through apoptosis (Figure 1.1B) (Goodman and Doe 1993; Schmidt *et al.* 1997; Lee and Luo 1999; Schmid *et al.* 1999; Karcavich and Doe 2005; Truman *et al.* 2010). These two progeny cells of a GMC often adopt different identities through binary cell specification, a mechanisms mediated by the Notch/Numb signaling pathway. The Notch/Numb pathway is used in various developmental contexts across diverse species (Artavanis-Tsakonas and Muskavitch 2010; Pierfelice *et al.* 2011). Here, its role in binary cell fate decision is observed in embryonic CNS lineages (Spana *et al.* 1995; Spana and Doe 1996; Buescher *et al.* 1998; Skeath and Doe 1998; Lundell *et al.* 2003; Karcavich and Doe 2005), as well as post-embryonic antennal lobe lineages (Lin *et al.* 2010; Lin *et al.* 2012), four other *engrailed* expressing brain lineages (Kumar *et al.* 2009), and most of the VNC lineages (Truman *et al.* 2010).

Recently, a different proliferation pattern was discovered. 8 brain NBs can produce transient amplifying precursors (TAPs, or named as intermediate neural progenitors, INPs), which can self-renew for several rounds and generate a short series of GMCs to deposit neurons and glia (Bello *et al.* 2008; Boone and Doe 2008; Bowman *et al.* 2008). They are named type II (in contrast to the general type I lineages) or PAN (posterior Asense-negative) lineages. The discovery of

this mammalian type of proliferation pattern in *Drosophila* brains makes studies of type II lineage an actively pursued subject for insights into CNS development in higher organisms (Izergina *et al.* 2009; Lui *et al.* 2011; Jiang and Reichert 2012; Bayraktar and Doe 2013; Viktorin *et al.* 2013; Wang *et al.* 2014)

Neurons of a specific NB lineage mostly do not migrate and reside together with their neurites bundling and extending through common tracts (Pereanu and Hartenstein 2006; Yu *et al.* 2013a), a feature that works as an advantage to study insect brains. Such elaborate and distinct morphologies of NB lineages are undoubtedly sensitive and faithful readouts of their own unique identities because structure, function, and cell fates are tightly linked in CNS development. Numerous genetic techniques have been developed to trace and study *Drosophila* nervous systems with valuable pedigree information (Lee and Luo 1999; Lai and Lee 2006; Yu *et al.* 2009; Awasaki *et al.* 2014). The effort to describe lineages in their original wild-type state will eventually pay off when new technologies allow us to further introduce perturbations for mechanistic cell fate/lineage identity studies.

Studying brain development requires efficient gene targeting technologies

Quite often, biology advances after acquiring valuable “drivers” to introduce reporters or effectors in certain cells of interests so that one can follow or manipulate them. To make use of Urbach and Technau’s embryonic NB map and correlate marker gene expression with succeeding NB development, I decided to

generate T2A-Gal4 knock-ins (KIs) for these molecular markers (Brand and Perrimon 1993; Szymczak *et al.* 2004; Diao and White 2012). Ideally, such drivers should capture the endogenous expression profiles of the targeted genes. Subsequently, with the help of a pan-NB specific *deadpan* (*dpn*) promoter plus several genetic designs, their expression can be further restricted to NBs that previously expressed the corresponding genes during development (Awasaki *et al.* 2014). When this intersection occurs in stem cells like NBs and triggers additional irreversible recombination events, all of the progeny cells can then be uniformly labeled by reporters or influenced by effectors (James Truman lab's termed this genetic strategy as "immortalization" for easy reference, and we adopted it in Awasaki *et al*. 2014).

Therefore, the possibilities exist to understand the cause-and-effect relationships between the molecular markers and their corresponding lineages once we immortalize and visualize them as a first step. However, to fully investigate the involvement of the surveyed 34 genes by the Technau group requires an efficient way to create related KI drivers, not to mention numerous other candidate genes that might play a role in this process. Hence, we recognized the urgent need to improve the current gene targeting (GT) technologies for *D. melanogaster*.

Existing strategies for efficient gene targeting

Genome engineering was pioneered by Dr. Capecchi and others in the 1980s (Capecchi 2005). Several lessons were learned over years of trial-and-error experiments done with cultured mouse embryonic stem (ES) cells. By artificially introducing a DNA template to ES cells, microinjection or electroporation, there is an about ~1/1000 chance that the endogenous target locus is perfectly modified with the template through a DNA repair mechanism called homologous recombination (HR). HR efficiency is cell-cycle dependent, showing a peak of activity during early S-phase. Also, linear DNA template is the preferred substrate for HR, and injecting a single copy of it is sufficient for HR. While HR is achievable through arduous work, non-homologous recombination (random integration) hampers the effort by its much higher rate of occurrence (over 100 fold). Fortunately, by implementing a negative selection module outside of the homology arms, such undesirable events can be easily removed.

Despite the success of the mouse story, such a system is not readily applicable for many model organisms because of their lack of culturable stem cell or germ cell system, which allow easy manipulation and direct germline transmission. The relatively low efficiency of HR by mere DNA template introduction discouraged most researchers from adopting gene targeting for their favorite organisms. However, with the advent of sequence-specific nucleases, genome engineering has become a more approachable technology and hopefully will evolve into a routine practice (Kim and Kim 2014). These programmable nucleases can create targeted double-strand breaks (DSBs) that result in small

indels (insertions or deletions) by error-prone non-homologous end-joining (NHEJ) reactions. Otherwise, with an appropriate DNA template, these DSBs can boost HR efficiency by 2-3 orders of magnitude (Rouet *et al.* 1994; Jasin 1996). For *Drosophila*, such enhanced efficiency consequently makes direct embryo injection a feasible GT strategy to introduce essential targeting materials into embryonic pole cells and collect targeted flies with limited amounts of effort. Hence, tailored nucleases has critically revolutionized the *Drosophila* GT technologies ovr the past few years.

The first nuclease of this kind is the Zinc-finger nuclease (ZFN) (Kim *et al.* 1996). It has been introduced and heavily adopted in *Drosophila* by the Carroll group (Bibikova *et al.* 2002; Bibikova *et al.* 2003; Carroll *et al.* 2006; Beumer *et al.* 2008). ZFNs are hybrid proteins composed of a DNA-binding domain of Cys_2His_2 zinc finger proteins and a nonspecific cleavage domain of the type IIS restriction enzyme *Fok*I. The cleavage domain must dimerize to cut DNA (Bitinaite *et al.* 1998; Smith *et al.* 2000). Such a prerequisite doubles the length of recognition sites, which substantially increases the specificity of ZFNs. Each zinc finger contacts three base pairs (bps) of DNA, and consecutive fingers bind continuous DNA triplets. To assemble a functional ZFN with proper specificity, each monomer contains 3-6 zinc fingers, which recognize 9-18 bps. Together, two monomers bind their own half sites that are separated by 5-7 bps (optimal 6 bp) and make staggered breaks and leave four-nucleotide 5’ overhangs.

Although ZFN seemed a promising solution to genome engineering when proposed, there exist several major setbacks in practic (Kim *et al.* 2010). First, not all of the 64 possible DNA triplets have corresponding pre-characterized zinc fingers, with 5'-GNN-3' being the most extensively tested set. This significantly hampers the design flexibility due to poor genome targeting density. Also, even though zinc finger-DNA interaction is believed to be modular in nature, not every combination of zinc fingers guarantees DNA targeting activity. Further, there is an imperfect correlation between ZFN activity *in vitro* and *in vivo* in *Drosophila*. Finally, the major disadvantage of ZFN is its possible off-target toxicity. This undesirable side effect likely comes from the fact that wild-type *Fok*I nuclease domain can still form a functional homodimers to cleave DNA when only one of the dimers binds its recognition site.

The enthusiasm for ZFN was gradually taken over by transcription activator-like effector nuclease (TALEN) (Boch *et al.* 2009; Moscou and Bogdanove 2009; Miller *et al.* 2011). TALEN is fundamentally similar to ZFN in terms of functional design, with both of them carrying the *Fok*I nuclease domain at their carboxyl termini. However, for sequence recognition, TALEN uses transcription activator-like effectors (TALEs), which are DNA-binding domains derived from the plant pathogenic *Xanthomonas* spp. bacterium. These domains contain tandem arrays of 33-35 amino acid repeats that in turn recognize sequential base pairs in the major groove of the target DNA. The specificity of the repeats comes from adjacent amino acids at position 12 and 13, which are

termed as repeat variable di-residues (RVDs). Currently, the four widely used RVDs used to recognize adenine, guanine, cytosine and thymine are Asn-Asn, Asn-Ile, His-Asp and Asn-Gly, respectively.

With the simple one-to-one correspondence between the RVDs and the four bases, one can easily design TALENs to target any DNA sequence, which is a significant advantage of TALEN over other tailored nucleases. Also, the easily increased specificity of TALENs leads to less off-target disruptions, and hence less cytotoxicity (Mussolino *et al.* 2014). However, constructing a DNA fragment for a TALEN is often challenging and time-consuming because the homologous sequences coding the repeats are prone to recombine with each other. Currently, researchers often outsource to companies or core facilities to acquire or design ZFNs and TALENs (Kim and Kim 2014).

Lastly, since its potential for targeted mutagenesis was shown (Jinek *et al.* 2012), the clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated (Cas) system has drastically transformed our ability to edit genomes of a variety of organisms. It is an evolved bacterial and archaeal adaptive immune defense against invading phages or foreign nucleic acids. Among the three types of CRISPR/Cas systems, type II is the simplest one to be adopted because it only requires three components to execute a RNA-guided cleavage: the target-specific CRISPR RNA (crRNA), the invariable target-independent trans-activating crRNA (tracrRNA), and the CRISPR-associated protein 9 (Cas9). Together, crRNA, tracrRNA, and Cas9 forms an active DNA

nuclease that recognizes and cuts a 23-bp target DNA, which consists of a 20-bp target site (i.e. protospacer) and then a 5'-NGG-3' sequence known as protospacer adjacent motif (PAM).

The reason CRISPR/Cas has quickly gained in popularity is its simplicity. First of all, functional features of crRNA and tracrRNA can be combined and expressed as a single guide RNA (gRNA), which further simplify type II CRISPR/Cas as a two-component system. More importantly, the specificity of CRISPR/Cas comes from the 20-bp target site, which can be easily generated by annealing of two short primers, and cloning them into a vector containing a gRNA backbone. Concerns about off-targets can also be addressed by using Nickase with double nicking (Cong et al. 2013; Ran et al. 2013). However, the targetable sequences of CRISPR/Cas are restricted by the requirement of a PAM sequence; also, they can be further limited by the requirement of a guanine at the 5' end if gRNAs are transcribed under a U6 promoter.

Golic+: an enhanced GT package for *Drosophila* CNS studies

Gene targeting by HR in *Drosophila* was achieved by Rong and Golic in 2000 (Rong and Golic 2000). However, this strategy has been deemed inefficient with few successful stories (Huang *et al.* 2008). To practice it, the template, termed donor DNA, is first inserted in the genome as a p-element, p{donor}. Then, linear donor DNA is provided stochastically in larval primordial germ cells through the actions of the heat shock induced flippase (FLP, circularization) and I-SceI

meganuclease (linearization). It is rare but possible that the released double-strand DNA template is used by the endogenous DNA repair machinery to edit the target locus by HR. Afterwards, this rare event is recovered by screening for red-eyed candidate targeted flies for carrying a mini-white eye marker, residing between the HR homology arms. With the low efficiency of unassisted HR and the amount of sorting work to collect the few GT candidates, this heat shock strategy pioneered by the Golic group unfortunately never gained the popularity its designers had wished for.

We intended to improve Golic GT technology by addressing the insufficiencies in its three major steps: generation of donor DNA, targeting, and screening. Briefly, a cystoblast-specific *bam* promoter was used to direct the expression of three proteins, FLP, I-SceI, and Cas9 in every cystoblast (CB) that are constantly produced in a female's ovary. FLP and I-SceI are responsible for the release of a pre-integrated donor DNA, while Cas9 works together with a universally expressed gRNA to generate a DSB at the target locus, and hence a much enhanced targeting efficiency. Finally, a lethality selection scheme was implemented to help with the recovery of GT candidates by eliminating non-targeted progeny.

This new GT package we developed is named Golic+ (an abbreviation for **g**ene targeting during **o**ogenesis with **l**ethality **i**nhibitor and **C**RISPR/Cas "plus") to honor the pioneering work of the Golic group with a "+" to indicate our significant improvements. Golic+ was built from scratch following insights from

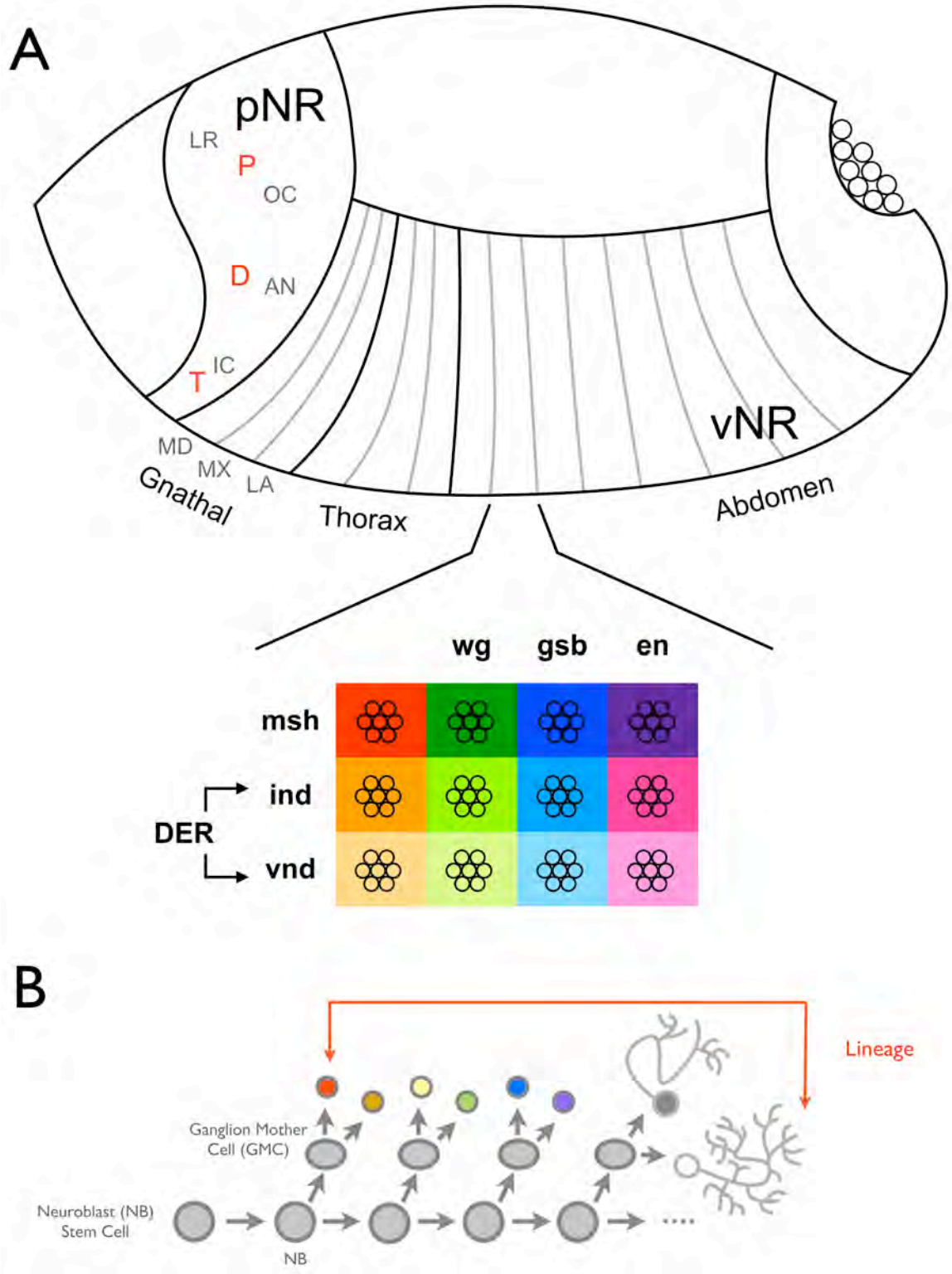
the mouse system plus a few add-ons from the fly community (Capecchi 2005; Huang *et al.* 2008). We cautiously tested the ideas and assembled the components step by step: visualizing effective provision of donor DNA in CBs, targeted mutagenesis in CBs with CRISPR/Cas, and identifying a robust repressor (miRNA)/toxic protein pair to implement the lethality selection. In the end, we generally observed a 10-fold increase of targeting efficiency while comparing Golic+ to traditional Golic heat shock strategy under close to identical genetic settings. Using *bam* promoter and our lethality selection scheme also brings us huge savings in time and labor. Quite often, 50% of the founder females produced correct targetings using Golic+.

Six molecular markers (*vnd*, *ind*, *msh*, *lab*, *unpg*, and *hkb*) were first selected to test our plan because of their characteristic concentrated expression in the relatively simpler tritocerebral and deutocerebral brain neuromeres. Six T2A-Gal4 KIs were hence generated, and they produced expression patterns that are identical to the ones of their corresponding genes. Through immortalization, we could follow the lineages made by NBs positive for a particular early patterning gene at larval and adult stages. Such information will guide our future molecular studies on NB lineage identities and their roles in neuronal diversification. Moreover, one can refine NB drivers using various intersection strategies based on the expression of early patterning genes, and thus target specific lineages for more detailed structure-functional and molecular studies.

Figure 1.1. Central nervous system development in *Drosophila*.

(A) Neuroblast patterning at embryonic stage. The roles of positional cues in NB determination are well-studied in *Drosophila* ventral nerve cord. There are two neurogenic regions on the ventral side of a fly embryo, the ventral neurogenic region (vNR) and the pro-cephalic neurogenic region (pNR). vNR gives rise to the ventral nerve cord, while the pNR forms the adult brain. Insects are segmented. Within each vNR segment, there exist AP (*wg*, *gsb*, and *en*) and DV (*msh*, *ind*, *vnd*, and *DER*) patterning axes that together create a grid-like Cartesian coordinate system. And each NB arise from a certain coordinate acquire its own unique spatial coding and hence unique identity. LR: labral; OC: ocular; AN: antennal; IC: intercalary; MD: mandibular; MX; maxillary; LA: labial; P: protocerebrum; D: deutocerebrum; T: tritocerebrum. Illustration is adapted from Technau *et al*., 2006. (B) Most of the neuroblasts in the brain undergo this proliferation pattern. They continually perform asymmetric cell divisions to renew themselves and produce intermediate progenitor cells called ganglion mother cells, which in turn divide one more time to produce two progeny cells, mostly neurons and sometimes glia. Together, all of the progeny cells from a specific NB are called a lineage.

Figure 1.1. Central nervous system development in *Drosophila*.



CHAPTER II

Developing a new efficient *Drosophila* gene targeting technology: Golic+.

1) This chapter has been published: Chen, H. M., Huang, Y., Pfeiffer, B. D., Yao, X., & Lee, T. (2015). An Enhanced Gene Targeting Toolkit for *Drosophila*: Golic+. Genetics, **199**: 683-694. Hui-Min Chen prepared the manuscript. Tzumin Lee wrote the paper.
2) This chapter has been modified to fit the style of the dissertation.
3) Tzumin Lee conceived the project. Yaling Huang, Barret Pfeiffer, and Xiaohao Yao provided molecular biology support for a subset of DNA constructs. Hui-Min Chen designed and conducted the experiments, collected and analyzed the data.

Introduction

Although the advent of genome sequencing has given us access to the code of diverse proteins, RNAs, and transcriptional regulatory elements, it gives no insight into how these sequences function. Unraveling the encrypted information requires a method to directly manipulate the genome in a controlled way. Ends-out gene targeting (GT) supplies just such a method by allowing seamless replacement of endogenous sequences with engineered DNA fragments (Thomas and Capecchi 1987). One can therefore place designer ‘genes’ into their native loci or otherwise edit the nucleotide sequences in any genomic region of interest.

GT depends on the occurrence of double-strand DNA crossing-over between the donor DNA and the genomic domain with which it shares extensive sequence homology. For ends-out GT, homologous recombination takes place between flanking homology arms of a linear donor DNA and a target genomic locus allowing insertion of an arbitrary DNA fragment into the endogenous target. However, homologous recombination occurs at low frequency and the linear donor DNA can be integrated into other genomic regions by non-specific insertion. Successful GT has therefore relied on effective strategies of recovering the rare GT events. To accomplish such screening requires selection markers both within as well as outside the homologous arms of the donor DNA (Capecchi 2005). However, false positives remain among those that have selectively

retained the internal marker. In mice where embryonic stem cells allow in-vitro GT, researchers can efficiently screen candidates to validate GT by DNA sequencing before proceeding through microinjection and the time-consuming labor-intensive mouse genetics (Capecchi 2005)

In the genetically powerful *Drosophila* where cultivatable stem cells are unavailable, Rong and Golic have pioneered a transgene-based system for implementing GT in the female germline (Rong and Golic 2000; Gong and Golic 2003) (Figure 2.1A). The Golic system avoids direct injection of the donor DNA into the embryonic primordial germ cells, which may be an untenable approach given the extremely low frequency of homologous recombination. Instead, it utilizes a P-element construct where donor DNA is flanked by *FRT* sites and contains a rare-cutting I-SceI site, which can therefore be transformed into a linear targeting molecule via the action of FLP and I-SceI. The linear donor DNA can be generated in the female germline by transient induction of *hs-FLP* and *hs-I-SceI* during larval development. This transient ubiquitous supply of the donor DNA at the early stage of germline development elicits GT in about 1% of female founders in a limited number of known cases (Huang *et al.* 2008). Although not systematically reported, it is known in the community that at some loci the GT rate drops significantly below 1%.

While the causes of reduced targeting efficiency are unknown, they may include inefficient release of the donor DNA (due to P element insertion site), *FRT* cassette size, characteristics of the homology arms, or the chromatin

‘context’ of the target loci. The challenges in scaling up the efforts to cope with rare GT events lie in both generating and screening for potential candidates. First, it requires heat-shock induction of FLP and I-SceI during the right stage of larval development to efficiently target the dynamic female germ (stem) cells. Second, it necessitates single-female crosses to ensure the independence of candidates as one adult female can yield multiple offspring with identical GT by the generation of multiple eggs from the same modified germline stem cell. Third, despite various improvements in the selection markers, it remains labor-intensive to screen out non-specific insertions from the large number of progeny generated in many independent crosses.

Recent advances in *Drosophila* transgenesis, binary transgene induction, site-specific recombination, repressible markers, and sequence-specific endonucleases have drastically expanded the arsenal for further improving the classical Golic system of ends-out GT (Venken and Bellen 2007; Kim and Kim 2014). For instance, transgenesis by phiC31-mediated site-specific integration can allow insertion of the donor DNA-containing construct at loci already proven to generate high-efficiency excision. With GAL4/UAS and LexA/lexAop (Brand and Perrimon 1993; Lai and Lee 2006; Pfeiffer *et al.* 2010), one can independently drive FLP/I-SceI to release the donor DNA in germ cells while executing the selection via targeted induction of other transgenes in different vital tissues. Therefore, a repressor can be incorporated into the donor DNA for recovering candidates based on suppression of toxic transgenes.

Moreover, the availability of programmable sequence-specific nucleases makes it possible to create DNA breaks at specific genomic loci of interest (e.g. use of zinc-finger nucleases in Bibikova et al. 2002, Bibikova et al. 2003, and Beumer et al. 2008). By provoking repair of the targeted locus, such methods can boost rates of gene disruption, editing, or addition by 2 to 3 orders of magnitude, levels that enable ready isolation of cells or organisms bearing a desired genetic change (Jasin 1996). The CRISPR/Cas system is particularly useful for various targeted genome modifications, as the Cas9 nuclease can be reliably directed by custom-made guide RNAs to make specific DNA cuts based on DNA-RNA base pairing (Jinek *et al.* 2012; Bassett *et al.* 2013; Cong *et al.* 2013; Gratz *et al.* 2013; Hwang *et al.* 2013; Kondo and Ueda 2013; Mali *et al.* 2013; Ren *et al.* 2013; Sebo *et al.* 2013; Yu *et al.* 2013b; Port *et al.* 2014). Within the last two years, CRISPR/Cas9 has been fruitfully applied in diverse species to create or correct mutations, insert transgenes at precise locations, and even generate knock-ins via targeted replacement of multi-kilobase genes (Harrison *et al.* 2014; Hsu *et al.* 2014; Kim and Kim 2014). The efficiency of CRISPR/Cas9-mediated genome editing, including knock-ins, is high and often exceeds 10% without selection (e.g. Byrne *et al.* 2014).

Ends-out GT by direct embryo injection thus became much more feasible (Baena-Lopez *et al.* 2013; Gratz *et al.* 2013; Port *et al.* 2014; Xue *et al.* 2014; Yu *et al.* 2014), but to scale up the microinjection is costly and to recover

independent targeting events demands collection of no more than one candidate from an injected organism (Figure 2.1B). As GT efficiency could vary drastically, injection may need to be scaled up to prohibitive levels when dealing with large inserts or difficult loci. We were therefore interested in establishing a GT-technology that is scalable and leverages the power of existing *Drosophila* tools to virtually guarantee targeting success, regardless of genomic loci or insert size.

Here we establish Golic+ (gene targeting during oogenesis with lethality inhibitor and CRISPR/Cas) as an enhanced transgene-based GT toolkit (Figure 2.1C). Multiple improvements are adopted to simplify, optimize, and ensure the independent production of the linear donor DNA in each serially derived cystoblast throughout female germline development. The assembly line of de-novo GT allows us to pool unsynchronized organisms and breed them in groups, without concern about resampling the same GT events. Notably, the CRISPR/Cas system is needed to promote homologous recombination in cystoblasts and thus realize GT during active oogenesis. Moreover, we have engineered a repressor-based pupal lethality selection such that only the strong candidates can eclose for further breeding and PCR validation of GT. In sum, Golic+ is built upon the existing genetic/transgenic platforms, but made to be scalable such that one fly lab can perform multiple GTs simultaneously at ease and with stronger chances of success.

Results

Producing donor DNAs in each cystoblast

Existing GT systems employ the relatively limited numbers of embryonic and larval germline stem cells that may give rise to multiple identical targeting events due to clonal expansion (Figure 2.1A,B). Here we explored the possibility of inducing GT in the serially derived cystoblasts (CBs) of adult ovaries, which each develop into a female germ cell and thus guarantee independent GT events among individual offspring (Figure 2.1C).

At the tip of each germarium in ovaries, there exist two to three germline stem cells (GSCs). These cells divide alternately, self-renewing themselves while generating a series of CBs (Spradling 1993). A CB-specific promoter has been isolated from *bam*, a gene essential for female germ cell differentiation (Chen and McKearin 2003). We tested the available *bamP-GAL4* with *UAS-GFP* carrying p10 terminator (Chen and McKearin 2003; Pfeiffer *et al.* 2012) and confirmed its selective expression in newborn CBs but not the GSCs (Figure 2.2A).

We next examined the ability of *bamP-GAL4* to drive *UAS-FLP* and mediate the excision of an *FRT* cassette from the genome of each CB, mimicking the process of donor DNA liberation. We made Pfife (p10-facilitated indicators of flip excision), a dual-reporter flip-out construct (Figure 2.2B). Pfife carries an *FRT* cassette, excision of which results in fusion of UAS with GFP on the residual

transgene and concomitantly reconstitutes *UAS-tdTomato* in the circularized *FRT* cassette. Thus we detect both the residual Pfife based on GFP expression and the derived extra-chromosomal cassette based on tdTomato expression. We induced flip-out using either *bamP-GAL4* or *nosP-GAL4* in adult ovaries (Figure 2.2C). *bamP-GAL4* elicited co-expression of GFP and tdTomato in newborn CBs of most, if not all, germaria. But *nosP-GAL4* only led to GFP expression in both GSCs and CBs. The earlier onset of *nosP-GAL4* in GSCs has apparently led to complete loss of *FRT* cassettes prior to adult oogenesis. By contrast, *bamP-GAL4* had maintained the FRT cassette in GSCs while promptly releasing it in newborn CBs, meeting our need for de-novo production of 'donor DNA' through adult oogenesis.

We further explored the possibility of placing donor DNAs at specific attP sites using the phiC31 integration system, to facilitate the initial transgenesis and eliminate the uncertainties in the efficiency of flip-out that varies with insertion sites. To match future donor DNAs in size, we enlarged Pfife to create a version, BPfife, which carries a 9kb *FRT* cassette. We determined the flip-out efficiency of BPfife in CBs for various attP sites (Figure 2.2D). After comparison, *attP40* on the second chromosome and *VK00027* on the third chromosome were identified as ideal sites for donor DNA integration.

Targeted mutagenesis by CRISPR/Cas in cystoblasts

Homologous recombination frequency can be greatly increased by double-strand DNA breaks created at specific loci by sequence-specific DNA nuclease, such as the CRISPR/Cas system (Cong *et al.* 2013; Mali *et al.* 2013), which utilizes a guide RNA (gRNA) to specify Cas9's cutting site. We next explored how to implement CRISPR/Cas in *Drosophila* CBs.

We cloned two endogenous *Drosophila U6* promoters, *dU6-2* and *dU6-3*, to drive gRNAs as transgenes (Wakiyama *et al.* 2005; Hernandez *et al.* 2007). Two target sites within the *yellow* body color gene were selected. The corresponding primers were annealed and inserted in front of the gRNA scaffold (Figure 2.3A) (Jinek *et al.* 2012; Cong *et al.* 2013; Mali *et al.* 2013), to derive four gRNAs using either *dU6-2* or *dU6-3*. To test the effectiveness of these gRNA transgenes inserted at the same *attP* site, we first crossed females with gRNAs against *yellow* to males carrying Cas9 expressed under the *act5C* promoter (*act5C-Cas9*) (Port *et al.* 2014). We assayed the change in offspring's body color from brown to yellow and detected allele-dependent penetrance of yellow body color phenotypes (Figure 2.3B). *dU6-3*-driven #1 gRNA gave the strongest phenotype with nearly complete yellow body color in all offspring. *dU6-2*-driven #2 gRNA, by contrast, exerted no observable effect on wild-type body color. The other two combinations produced intermediate phenotypes with yellow patches in various mosaic patterns. These results indicate that *dU6-3* is a stronger promoter than *dU6-2* in the induction of gRNAs and that different target sequences vary in their efficiency of CRISPR/Cas-mediated mutagenesis.

Next we used *bamP-GAL4* to drive expression of either wild type or D10A nickase Cas9 (Cong *et al.* 2013), to induce *yellow* mutations guided by the above four gRNAs in CBs. 10 founder *y*+ females of each genotype were crossed to y^1 males, and we scored the occurrence of F_1 male yellow progeny (Figure 2.3C). We recovered in total only 10 *y*- male offspring, seven of which arose from the *UAS-Cas9*/*dU6-3-gRNA-y#1* pair, indicating again that *dU6-3* outperformed *dU6-2* and gRNA-y#1 was more efficient than gRNA-y#2. We further sequenced the genomic regions around the target site #1 and #2, to map the induced *yellow* mutations. We were surprised to observe double reads of *yellow* in eight of the ten *y*- males that supposedly carry only one *yellow* gene on X chromosome. We subsequently realized that the *attP2* site used for integration of UAS-Cas9 and various gRNAs transgenes was marked with *y*+. Due to the presence of two *yellow* alleles that presumably differed due to insertions and/or deletions, the reads from these samples overlapped and were not clearly interpretable. We did, however, have two samples, presumably in cases where a large deletion in one copy of *yellow* removed the sequencing primer, where clean reads showed small deletions around the expected Cas9 cleavage sites. Taken together, the *bamP-GAL4*-dependent induction of Cas9 can potently disrupt two copies of a wild-type gene in one generation, demonstrating the efficacy of CRISPR/Cas in manipulating CB genomes. But the need for mutating two copies of *yellow* at the same time may have prevented us from detecting the activity of the apparently weaker D10A nickase. In sum, we have established an effective strategy for

implementing CRISPR/Cas specifically in CBs also using the same *bamP-GAL4* driver such that it can be paired with production of donor DNA.

Repressor-based lethality selection for efficient GT screening

Having independent trials of GT among CBs that each form a mature oocyte allows group culture of many founder female flies in one bottle without concerns of redundantly sampling the same targeting event. Our next goal is to enrich for the correct targeting event with a crossover at each homology arm. We envisioned a repressor-based lethality selection to eliminate offspring that no longer carry the donor DNA or retain the entire donor DNA either at the original site or at a new site via non-specific insertion. This repressor or the co-induced toxic genes (see below) should be traceable during transgenesis and can replace the commonly used, but rather large mini-white marker.

Fulfilling the above conditions requires a repressible toxic gene as well as a non-repressible toxic gene, which can be induced to give visible phenotypes or cause pupal lethality depending on drivers. We chose $Rac1^{V12}$, encoding a constitutive active small GTPase, as the toxic gene (Luo *et al.* 1994). We made $Rac1^{V12}$ repressible by inserting at its 5' untranslated region the target sequences of a proven potent transgenic miRNA against rat rCD2, establishing the rCD2 miRNA (rCD2i) as the repressor (Chen *et al.* 2007; Yu *et al.* 2009). We placed both the repressible riTS-$Rac1^{V12}$ and non-repressible $Rac1^{V12}$ transgenes as well as rCD2i under the control of lexAop promoters of various strengths (Lai and

Lee 2006; Pfeiffer *et al.* 2010). We explored the use of eye-specific *GMR3-LexA::GADd* and pan-neuronal *nSyb-LexA::p65* for inducing visible phenotypes and pupal lethality, respectively. We settled on the weakened 5XLexAop2 for the control of either repressible or non-repressible Rac1^{V12} while reserving the original lexAop (Lai and Lee 2006) to drive rCD2i strongly. In that combination, we could consistently elicit a rough bar eye phenotype with *GMR>Rac1^{V12}* and observe numerous dead pupae without escapers in the presence of *nSyb>Rac1^{V12}* (Figure 2.4A). Both phenotypes were completely suppressed when rCD2i was co-induced with repressible Rac1^{V12}.

We assembled a GT backbone (pTL1, targeting with lethality selection) from scratch by adding the above transgenes in specific orientations plus various features, including *FRT/loxP* sequences, I-SceI/I-CreI/I-CeuI cutting sites, an attPX integration site, and multiple cloning sites for 5' & 3' homology arms, into a basic plasmid (Figure 2.4B). Briefly, *lexAop-rCD2i*, flanked by direct repeats of *loxP*, resides between the multiple cloning sites for 5' and 3' homology arm insertion, and the non-repressible *5XLexAop2-Rac1^{V12}* sits in an opposite orientation following the 3' cloning sites. In addition, another *5XLexAop2* promoter and the repressible Rac1^{V12} sequence lie before and after the first and second *FRTs* respectively such that upon flip-out, a *5XLexAop2-FRT-riTS-Rac1^{V12}* module is reconstituted to serve as the inducible and suppressible toxicity background. Finally, the I-SceI cutting sites are positioned next to *FRTs* for linearizing circularized donor DNA, and two rare-cutting sites (I-CreI and I-

CeuI sites) plus an attPX integration site are included for potential retargeting (Huang *et al.* 2011).

Carrying a non-repressible *5XLexAop2-Rac1^{V12}* in the *FRT* cassette outside the recombination region permits tracking of the original transgene based on the bar eye phenotype induced by *GMR3-LexA::GADd*. It further allows us to kill those GT offspring, which fail to lose the non-repressible Rac^{V12} due to no excision or non-specific insertion, at the pupal stage with *nSyb-LexA::p65*. Among those that have lost the non-repressible Rac1^{V12} following excision of the *FRT* cassette, a repressible *5XLexAop2-riTS-Rac1^{V12}* is automatically reconstituted and only the offspring with an insertion of the repressor, *lexAop-rCD2i*, contained within the 5' and 3' homology arms could survive. Viable candidates likely represent correct GT given loss of the non-repressor Rac1^{V12} and presence of the repressor transgene (Figure 2.4C).

High-efficiency ends-out gene targeting in cystoblasts

To validate our multifunctional GT backbone (pTL1), determine the stringency of the repressor-based lethality selection, and ultimately test GT in CBs, we went on and built a donor DNA construct for deleting *HES-related* (*Her*) by ends-out GT. *Her* resides on X chromosome and encodes a putative basic helix-loop-helix (bHLH) transcription factor with no reported function, possibly due to lack of null alleles.

Following integration of *Her* donor DNA at *attP40* on second chromosome as revealed by rough bar eyes in the presence of *GMR3-LexA::GADd*, we first tried GT using *hs-FLP* and *hs-I-SceI* to excise the donor DNA during early or late larval development as in conventional Golic systems (Figure 2.5A). ~100 founder females were individually crossed with n*Syb-LexA::p65* males per induction scheme with heat shocks at early or late larval stages (Table 2.1). All but eight pupae, derived from six founder females, died before eclosion. The survivors were subject to genetic mapping followed by genomic PCR confirmation (Figure 2.6A and 2.9C). Correct GT with deletion of *Her* was confirmed in five of the eight candidates, but the rCD2i repressor was absent in the other three survivors, which lacked proper targeting. Those rare escapers showed various defects in the reconstitution of the repressible *5XLexAop2-FRT-Rac1*^{V12} at *attP40* during excision of the donor DNA FRT cassette, explaining why the repressor was dispensable for their survival (Figure 2.6C). These results validated our multifunctional GT backbone, including the built-in repressor-based pupal lethality selection.

We next tried GT of *Her* in CBs by excising the same donor DNA from *attP40* via *bamP-GAL4*-dependent induction of FLP and I-SceI (Figure 2.5B). We found survivors in 37 of the 180 single-female crosses, but failed to identify any correct GT events (Table 2.1). 24 crosses yielded escapers whose survival no longer depends on rCD2i. The remaining false positives recovered from 13 vials had lost the non-repressible 5XLexAop2-Rac1^{V12} but maintained lexAop-rCD2i

on the original second chromosome (Figure 2.6C). The recovery of many more escapers together with the first seen non-specific insertions indicates that despite generation of excised donor DNA, homologous recombination is less efficient in CBs than primordial germ cells.

We have demonstrated above that the same *bamP-GAL4*-dependent CB-specific induction of Cas9 could effectively disrupt the *yellow* gene at the gRNA recognition sites. Creating sequence-specific DNA breaks by CRISPR/Cas may promote homologous recombination and thus realize GT in CBs. To test this, we made various *Her*-specific gRNAs and introduced them with *UAS-Cas9* into GT founder females (Figure 2.5C). We were excited to recover offspring with correct GT from over 10% of founder females that carried *dU6-3-gRNA-Her#1* (Table 2.1). Correct GT could be also found in about 4% of founder females with *dU6-3-gRNA-Her#2*. We have even uncovered one correct GT utilizing the D10 nickase, which was less efficient in our previous experiments. Notably, CRISPR/Cas enabled GT in CBs without increasing incorrect targeting events. The CRISPR/Cas-mediated GT, like targeted mutagenesis by CRISPR/Cas, depends on the strength of both Cas9 and the gRNAs, as the wild-type Cas9 and *dU6-3* consistently outperform the D10 nickase Cas9 and *dU6-2*, respectively. In addition, CRISPR/Cas shows variable efficiency based on the gRNA target site.

To extend the generality of the above conclusions, we performed a second ends-out GT using pTL1. We targeted the *HLHm5* gene on third chromosome that, like *Her*, encodes a putative bHLH transcription factor and

lacks known mutations. A donor DNA for deleting *HLHm5* by homologous recombination was inserted at *attP40* (Figure 2.6B). We further made two *dU6-3-gRNAs* targeting different *HLHm5* sequences. By conventional heat shock induction of GT in larval primordial germ cells, we obtained correct GT in about 1% of founder females (Table 2.1). When the same donor DNA was released specifically in CBs where Cas9 was co-induced to act with *HLHm5* gRNA, GT was strongly boosted: about 50% of founder females have yielded offspring with correct GT and multiple, supposedly independent, events could readily occur in one founder female (Table 2.1). Importantly, the percentage of false positives, including escapers and non-specific insertions, to founder females remained around 30% despite the high efficiency of GT. With these potent gRNAs, the D10 nickase Cas9 could effectively elicit GT in CBs as well, though at about one-tenth of the frequency as compared to the wild-type Cas9.

Golic+: a toolkit for *D. melanogaster* HR gene targeting

In order to make “GT in CBs” readily applicable, we embarked on simplifying the genetic workflow and number of transgenes. First, we incorporated the gRNA backbone into the GT plasmid and made pTL2 that allows integration of donor DNA and gRNA into the same attP site via one transgenesis (Figure 2.4B). We also removed one of the two I-SceI cutting sites to reduce the chance of premature I-SceI cutting and hopefully prevent those genomic aberrations noted in escapers. Second, we adopted the 2A peptide

technology to express all three enzymes (Cas9, FLP, and I-SceI) in one transcript under the control of either UAS or two adequate *bam* promoters of different lengths, bamP198 and bamP898 (Chen and McKearin 2003; Szymczak *et al.* 2004; Diao and White 2012). To simplify the genetic crosses, all common transgenes required (Table 2.3) were purposefully pre-assembled in two of the three major chromosomes so that the target chromosome is left untouched (Table 2.4).

To test these additional modifications, validate our design of Golic+, and demonstrate its efficiency in ends-out GT, we explored the possibility of knocking T2A-Gal4 into two spatial patterning genes: *msh*, the dorsal columnar gene in the ventral neuroectoderm (Isshiki *et al.* 1997), and *runt*, a pair-rule class segmentation gene (Kania *et al.* 1990). We engineered the *msh*-T2A-GAL4 and *runt*-T2A-Gal4 donors as well as corresponding gRNAs targeting the 3' ends of *msh* and *runt* open reading frames (Figure 2.7A,C) (Diao and White 2012). The *msh*-T2A-GAL4 donor DNA contains 3kb homologous arms, while the *runt*-T2A-GAL4 donor DNA carries 2kb ones. They were both integrated at *attP40* on the second chromosome, for targeting *msh* residing on the third and *runt* on the X chromosomes, respectively. Founder females inheriting *bamP-Cas9-2A-FLP-2A-I-SceI* (abbreviated as *bamP-CFI*) or *bamP-Gal4* + *UAS-Cas9-2A-FLP-2A-I-SceI* (abbreviated as *UAS-CFI*), plus a homozygous inducible and suppressible toxic background ({donor, gRNA}* / *5X-riTS-Rac1^{V12}*) were mated with male *nSyb-LexA::p65* for lethality selection (Figure 2.5D).

We could recover offspring with correct *msh-T2A-Gal4* targetings from all three versions of Golic+, two kinds of direct induction of CFI (under *bamP198* or *bamP898*) and one Gal4/UAS binary induction (Table 2.2). Both *bamP898-CFI* and "*bamP-Gal4* + *UAS-CFI*" performed at high GT rates around 60% (18/30, and 53/90, respectively in terms of "# of correct GTs / # of founder females"). However, they also resulted in frequent occurrences of false positives. In the case of *bamP198-CFI*, its GT rate was reasonably high (38%, 36/95) with a lower rate of false positives. Additionally, we observe about 40 - 50% of founder females yielded unhealthy candidates that died within days after eclosion. Correct *msh-T2A-Gal4* KIs were validated by sequencing of PCR products as well as visualization of GAL4 activity with GFP reporter in embryos (Figure 2.7A,B).

bamP198-CFI was chosen as the preferred Golic+ reagent to perform *runt-T2A-Gal4* targeting because it resulted in higher proportion of correct *msh-T2A-Gal4* GTs among the candidates. We carried out a group culture for 50 founder females and recovered 35 viable adults in total (Table 2.2). Subsequent genetic mapping revealed 12 candidates carrying lexAop-rCD2i on the targeted X chromosome. Genomic PCR and GAL4-induced embryonic expression confirmed all these 12 candidates were correct *runt* KIs (Figure 2.7C,D). Notably, all of the correct *runt* KIs were initially recovered as heterozygous females and later found the KI allele to be homozygous as well as hemizygous lethal. This phenomenon indicates that knocking in T2A-GAL4 somehow impairs

endogenous *runt* function or drives GAL4 expression in a toxic pattern or level. Moreover, it implies that a comparable number of hemizygous male candidates that carry correct runt knock-ins should have existed but died precociously during the initial lethality screen. Taking this into consideration, Golic+ has achieved a consistent efficiency in the ends-out GT at the success rate of about recovering 50, likely independent, correct GTs from 100 founder females.

Taken together, we have revolutionized the conventional Golic system of GT in three aspects that include automatic independent induction of GT in CBs, promotion of GT with CRISPR/Cas, and recovery of correct GT based on suppression of pupal lethality (Figure 2.8). We term this enhanced GT system as Golic+, abbreviated from ‘Gene targeting during oogenesis with lethality inhibitor and CRISPR/Cas’ as well as indicate our improvements to the pioneering Golic GT system. {donor, gRNA} transformants can be easily identified by the rough-eyed phenotype induced by GMR3-LexA (Firgure 2.9A). All common transgenes necessary for Golic+ are pre-assembled on two of the three major chromosomes, leaving the target chromosome empty (Figure 2.9B; Table 2.4).

Discussion

In the original and all previous modified Golic systems, the donor DNA is transiently released by heat shock and subsequently lost in most, if not all, cells at the mid-larval stage when the developing ovaries carry female germline stem cells plus a limited number of cystoblasts (CBs). By contrast, Golic+ ensures a continuous supply of the linear donor DNA to the serially derived CBs throughout female reproduction. This creates an assembly line with each newborn CB experiencing an independent trial of gene targeting (GT). Given the independent nature of GT occurring in CBs, we can further pool unsynchronized Golic+ founder females and breed them together in one bottle per GT. Golic+ thus converts the once highly involved and often unpredictable GT process into standard straightforward genetic crosses, such that one can perform multiple GT experiments simultaneously. If needed, one can scale crosses up indefinitely to recover targeting events even at loci with significantly lower targeting efficiency.

Fortunately, the *bam* promoter permits appropriately timed induction of transgenes selectively in newborn CBs. We can therefore maintain the resident donor DNA in the female germline stem cells and then excise it only in the serially derived CBs. Using a reporter construct mimicking the donor DNA in both length and the arrangement of *FRTs* and I-SceI site, we have identified the attP sites where donor DNA can be efficiently flipped out. We have further optimized

the *bam* promoter-dependent induction to maximize the percentage of CBs that have received the linear donor DNA during active oogenesis.

Golic+ further employs a repressor-based pupal lethality selection to facilitate the recovery of potential GT events. Our optimization in the supply of linear donor DNAs to CBs has led to the recovery of many more false-positive offspring, as compared to the conventional mid-larval pulse induction. False positives appeared at a rather constant level with rates compared to founder females being around 30%, but the frequencies of correct GT could be drastically increased with potent Cas9 and gRNA transgenes. The promotion of correct GT by CRISPR/Cas did not reduce false positives, suggesting independent sources of escapers, non-specific insertion, and correct GT. We suspect that non-specific insertion, which all reside on the original chromosome, arose by a common mechanism involving local rearrangement or hopping (Figure 2.6C). This would imply that the correct GT events enabled by CRISPR/Cas are derived through recruitment of an otherwise lost pool of liberated donor DNAs. To achieve GT in CBs absolutely requires CRISPR/Cas9, but comparable efforts could accomplish GT using mid-larval pulse induction without CRISPR/Cas. This phenomenon further suggests that the efficiency in capture of the liberated donor DNAs by homologous recombination is cell type or cell cycle-dependent.

The recent introduction of CRISPR/Cas has made GT via direct injection of donor DNAs plus supporting reagents (e.g. guide RNA) into early embryos possible in *Drosophila* (Baena-Lopez *et al.* 2013; Gratz *et al.* 2013; Bassett and

Liu 2014; Port *et al.* 2014; Xue *et al.* 2014; Yu *et al.* 2014). In addition to rapid turnover, direct embryo injection allows easy adoption of diverse GT strategies. Can Golic+ secure the female germline as the preferred site for GT in *Drosophila melanogaster*? First, it is more efficient and scalable to generate independent trials of genome modification in the continually generated CBs than within the fixed/small pool of embryonic primordial germ cells. Second, the reliability of direct injection remains unclear. Can one consistently recover correct GT from an affordable scale of embryo injection? As large inserts or difficult loci may reduce GT efficiency by orders of magnitude, injection may need to be scaled up to prohibitive levels. To repeat microinjections with freshly prepared DNAs and RNAs is further costly, labor-intensive, and time-consuming. In addition, the great scalability of Golic+ may allow further reduction in the length of homology arms required for efficient GT (Beumer *et al.* 2013).

Despite the current success in Golic+, several issues remain to be addressed. First, to eliminate escapers, we have deliberately labeled *5X-riTS-Rac1^{V12}* with *3xP3-RFP*, hoping that we can pre-select for red-fluorescent-eyed candidates that must carry the repressor, *lexAop-rCD2i*. Yet, candidates with 3xP3-RFP were a minority among correctly targeted candidates, possibly because residual donor (*5X-FRT-riTS-Rac1^{V12}*) is less toxic than *5X-riTS-Rac1^{V12}* with the extra *FRT* sequence, and hence the bias. To reduce this bias, we will lessen the selection toxicity from *5XLexAop2-* to *3XLexAop2-riTS-Rac1^{V12}*. Second, we notice that non-specific insertion was not reported with direct embryo

injection. We suspect that using circular versus linear donor DNA might underlie this stunning difference. Additionally, it has been argued that circular donor DNA outperforms the linearized form in the targeting efficiency (Beumer *et al.* 2008; Gratz *et al.* 2014). Hence, we will generate and explore the effectiveness of *bamP-Cas9-2A-FLP* in eliciting GT in CBs. Finally, while several features (I-CreI, I-CeuI, and attPX) have been built in for re-targeting, recombinase-mediated cassette exchange (RMCE; Schlake and Bode 1994) appears to be a superior approach. We will therefore add the ability for RMCE (recombinase-mediated cassette exchage) into our next version of Golic+.

In sum, using the widely available phiC31 integration system one can reliably insert donor DNA and corresponding guide RNA, in one construct, into pre-tested attP sites. The remaining transgenes are then supplied via common fly stocks support (Table 2.4): (i) detecting and balancing the initial transformants (*GMR3-LexA::GADd*), (ii) conducting cystoblast-specific GT (*bamP-Cas9-2A-FLP-2A-I-SceI* and *5XLexAop2-riTS-Rac1*^{V12}), and (iii) implementing the pupal-lethal selection (*nSyb-LexA::p65*). Only strong candidates can eclose and be PCR validated immediately after breeding. The entire procedure will take just two rounds of en-masse crosses after the establishment of the starter line carrying donor DNA and guide RNA (Figure 2.9B). Using only well-established genetic/transgenic techniques, the relatively effortless Golic+ should empower all fly labs to perform sophisticated ends-out GT.

Materials and Methods

DNA constructs

The constitutively active *Drosophila* Rac1^{V12} was created by amplifying the wild type Rac1 coding sequence (CDS, gift of Julian Ng) with a forward PCR primer containing the GGA to GTG mutation. Following rules in previous studies (Chen *et al.* 2007; Yu *et al.* 2009), 6 different rCD2 miRNAs were created and compared for their effectiveness in suppression. Here are their rCD2 target sequences:

#1: CATGGCATCAACCTGAACATCC / ACAAATGGGACACGTATCCTGG;
#2: GGTTGCCGAGTTTAAAAGGAAG / GCACTGGACTTGAGGATTCTAG;
#3: ACAGATGTTGAACTAAAGCTGT / GAAATCGGGAGCATTTGAGATC;
#4: GGAGACTTGAAGATAAAGAATC / CGTCAGAAGACCATGAGTTACC;
#5: GATGACAGTGGCACCTATAATG / TCAACTGTCCAGAGAAAGGTCT;
#6: GTAACGGTATACAGCACAAATG / GATAAAAGCTTCCAGAATGAGC;

rCD2 miRNA #4 and #6 had indistinguishable full suppression performance, and #6 was selected for it being the Chinese lucky number. 8XLexAop2-ri6TS-Rac1^{V12}-hsp70 was created by cloning ri6TS-Rac1^{V12} into pJFRC18 (*Not*I/*Xba*I) (Pfeiffer *et al.* 2010) and replacing the original SV40 terminator (SV40T) with hsp70 terminator (hsp70T, *Xba*I/*Fse*I). 5XLexAop2-ri6TS-Rac1^{V12}–hsp70T was generated subsequently by removing the 3XLexAop2 between the *Avr*II and *Nhe*I sites of 8XLexAop2-ri6TS-Rac1^{V12}-hsp70T (self-ligation with the *Avr*II/*Nhe*I compatible ends). We further created a 3xP3-RFP (instead of mini-white) version of the 5XLexAop2-ri6TS-Rac1V12-hsp70T transgene by replacing the original mini-white marker (*Asc*I, filled-in) with a 3xP3-RFP-SV40 fragment (Horn *et al.*

2000). Additionally, to reduce similarity and chances of recombination between sequences, we ordered a *Drosophila* codon-optimized coding sequence of Rac1^{V12} (GeneArt, invitrogen), and created the non-repressible 5XLexAop2-opRac1^{V12}–hsp70T (*Not*I/*Xba*I). rCD2 miRNA #6 was cloned into pMLH (*Not*I/*Xho*I) (Awasaki *et al.* 2014) to make lexAop-rCD2i.

pTL1, pTL2, Pfife, and BPfife were constructed following traditional molecular cloning and assembled from smaller DNA fragments. Their sequences and detailed annotations can be found on Addgene (https://www.addgene.org/) once deposited. To assemble pTL1, smaller DNA modules were added sequentially by attaching them either on forward or reverse PCR primers. First, a starting cassette (5XLexaop2-FRT-ISceI-FRT-ri6TS-Rac1^{V12}-hsp70T) was assembled from a triple ligation of two PCR fragments (*Hin*dIII-5XLexaop2-FRT-I-*Sce*I, I-*Sce*I-FRT-ri6TS-Rac1^{V12}-hsp70T-*Eco*RI), and a *Hin*dIII/*Eco*RI digested pBluescript II KS(+) plasmid. Secondly, after digesting this starting cassette with I-*Sce*I, two more PCR fragments (I-*Sce*I-5'MCS-I-*Cre*I-attPX-loxP-*Spe*I and SpeI-hsp70T-opRac1^{V12}-5XLexAop2-I-*Sce*I) were added in to create a bigger intermediate cassette (5XLexAop2-FRT-I-*Sce*I-5'MCS-I-*Cre*I-attPX-loxP-*Spe*I-hsp70T-opRac1^{V12}-5XLexAop2-I-*Sce*I-FRT-ri6TS-Rac1^{V12}-hsp70T). This intermediate cassette was then transferred (*EcoR*I and partial *Hin*dIII digestion) to pJFRC-MUH (*Hin*dIII/*Eco*RI) (Pfeiffer *et al.* 2010) after removing the mini-White marker in pJFRC-MUH with *Asc*I cutting and self-ligation. Lastly, pTL1 is finished by adding into the *Spe*I site the final PCR fragment *Spe*I-lexAop-rCD2i-

loxP-I-*Ceu*I-3'MCS-*Xba*I (*Spe*I and partial *Xba*I digestion), which was assembled from two smaller PCR fragments, *Spe*I-lexAop-rCD2i-*Pac*I and *Pac*I-loxP-I-*Ceu*I-3'MCS-*Xba*I.

To create pTL2, we first removed the I-*Sce*I site downstream of the 3' MCS of pTL1. pTL1 was partially digested with I-*Sce*I and re-circularized using Gibson assembly (New England BioLabs, Inc., Gibson Assembly Master Mix, E2611L) with a small PCR fragment containing one *Avr*II site at each end. This fragment was later removed by *Avr*II digestion and self-ligation so that the original I-*Sce*I site (TAGGGATAACAGGGTAAT) became TAGGGATAA-CCTAGG-ATAACAGGGTAAT. To add gRNA backbone, the plasmid was digested with *Eco*RI, filled-in, and then ligated with a *Hind*III/*Eco*RI-digested, filled-in dU6-3-gRNA fragment (see below and Figure 2.3A). The orientation with the *dU6-3* promoter lying adjacent to the suppressible ri6TS-Rac1^{V12} was selected in pTL2.

To assemble Pfife, a new FRT cassette (*Not*I-FRT-*Bgl*II-*Hin*dIII-*Fse*I-*Bam*HI-FRT-*Xho*I) was first created by annealing two partially overlapping primers, filling in both strands with PCR, and cloning it into pBluescript II KS(+) (*Not*I/*Xho*I). Afterwards, DNA fragments (*Hin*dIII-10XUAS-*Bgl*II, *Hin*dIII-BPStop-*Fse*I, and *Bam*HI-tdTomato-*Xba*I-p10-*Ngo*MIV) were sequentially added in to generate a FRT-10XUAS-BPStop-p10-tdTomato-FRT cassette. BPStop is a DNA module that by design should effectively repress/eliminate read-through of transcription and translation from nearby promoters (Sauer 1993; Zinyk *et al.*

1998). Finally, this larger FRT cassette was excised (*Not*I/*Xho*I) and put into pJFRC28 (Pfeiffer *et al.* 2012) to create Pfife. To further generate Bfife, we first inserted a ~5kb DNA fragment (from plasmid KB700, *Hin*dIII/*Not*I-digested, filled-in) into the pBluescript II KS(+)-FRT-10XUAS-BPStop-p10-tdTomato-FRT plasmid (*Spe*I digested, filled-in). Then, this ~9kb FRT cassette was put into pJFRC28 to generate BPfife (*Not*I/*Xho*I).

To create UASt-FLP in this study, we first assembled pJFRC-MUH-IVS-WPRE-hsp70T by replacing the original SV40T of pJFRC-MUH with hsp70T (*Xba*I/*Ngo*MIV) and further inserted *Bam*HI-IVS-*Bgl*II-*Not*I-*Xho*I-*Kpn*I-*Xba*I-WPRE-*Spe*I into the *Bgl*II/*Xba*I-digested pJFRC-MUH-hsp70T. Then, a *Bgl*II-FLP-*Xho*I PCR fragment was created using pBPopFLP(1GtoD)Uw as the template. Afterwards, *Xho*I-3XGST-securin N50 a.a.-XbaI was added on to the fragment with primers and PCR amplification. Finally, the whole *Bgl*II-FLP-*Xho*I-3XGST-securin N50 a.a.-*Xba*I fragment was inserted into pJFRC-MUH-IVS-WPRE-hsp70T (*Bgl*II/*Xba*I) to create UASt-FLP (UAS-IVS-opFLP(1GtoD)::3XGST::securing N50 a.a.-WPRE-hsp70T). The UASp-FLP used in this study actually contains UASp-FLP and UASp-I-SceI separated by an insulated spacer cassette (Pfeiffer *et al.* 2010). To assemble it, a blunt end/*Eco*RI-digested PCR product (*Avr*II-UASp promoter-MCS-K10 terminator-*Fse*I-*Pme*I-*Nhe*I-*Eco*RI) (Rørth 1998) was first inserted into pJFRC-MUH (*Hin*dIII, filled-in/*Eco*RI) to make pUASp-attB. Second, the *Kpn*I/*Hin*dIII filled-in fragment from pBPopFLP(1GtoD)Uw was inserted into pUASp-attB (*Kpn*I/*Xba*I filled-in) to

make pUASp-attB-FLP. Third, a synthetic *Drosophila* codon optimized I-SceI (GeneArt, invitrogen) was inserted into pUASp-attB (*Kpn*I/*Xba*I) to make pUASp-attB-I-SceI. Fourth, the insulated spacer cassette flanking with *Ngo*MIV sites was inserted into pUAS-pattB-I-SceI (*Ngo*MIV) to make pUASp-attB-I-SceI-insulator. Finally, the *Avr*II/*Pme*I fragment from pUASp-attB-FLP is inserted into *Nhe*I/*Pme*I site of pUASp-attB-I-SceI-insulator to make UASp-FLP (pUASp-attB-I-SceI-insulator-UASp-FLP).

To recover {donor} transgenes, we generated an eye-specific LexA driver, GMR3-LexA::GADd, which contains three copies of a truncated glass-binding site (Ellis *et al.* 1993; HAY *et al.* 1994) to drive LexA::p65 (Pfeiffer *et al.* 2010). nSyb-LexA::p65 is a neuronal synaptobrevin promoter-fusion LexA::p65 (Pfeiffer *et al.* 2010; Awasaki *et al.* 2014) construct created as the essential (neuronal) driver to implement the larval/pupal lethal selection of Golic+. *Drosophila* codon-optimized CDSs of FLP (opFLP(1GtoD), *Xho*I/*Xba*I), I-SceI (GeneArt, invitrogen, *Kpn*I/*Xba*I), and Cas9 (gift of Justin Crocker, *Kpn*/*Xba*I) were cloned into pJFRC28 for enhanced ovary expression (Pfeiffer *et al.* 2012). Additionally, Syn21 was added before start codons of optimized FLP and I-SceI to enhance their expression (Pfeiffer *et al.* 2012). For Cas9, we further added in its N-terminus a nuclear localization signal (CCAAAGAAAAAGAGAAAGGTT) in the hope of increasing its effectiveness. Cas9 D10A mutation was introduced by amplifying Cas9 with a forward primer containing a GAC to GCC codon change (Cong *et al.* 2013). To create bamP198-CFI and bamP898-CFI, *bam* 3'UTR was

cloned from BACR06L08 into pJFRC28 (*Kpn*I/*Eco*RI) with primers (GGGGTACCTCTAGACTAATGCTGTGCACATCGATAAAAG and GGAATTCAGTCCAAACACAAATCGTAAATATTTATTTG). bamP198 and bamP898 5'UTRs were then cloned (BACR10P10; primers: bamP198 forward: CCAAATCAGTGTGTATAATTGTAGTTAAAATG; bamP898 forward: AGATCTAACCATTGATTAACTTCACAAC; and common *bam* 5'UTR reverse: GCTCTAGAGGTACCTAAGTTAAATCACACAAATCACTCGATTTTTG) and added into the bam 3'UTR-carrying vector, respectively (*Sph*I digestion, blunt/*Xba*I). Finally, Cas9-P2A-FLP-E2A-I-SceI was assembled and added into the previous vector (*Kpn*I/*Xba*I) using Gibson Assembly. P2A and E2A were chosen among 2As for their higher cleavage efficiency (Kim *et al.* 2011). Also, Gly-Ser-Gly linkers were put in front of these 2As as suggested for improved cleavage efficiency (addendum to Szymczak *et al.* 2004).

dU6-2 and *dU6-3* promoters were cloned from BAC clone BACR47D16. Two *Sap*I sites were put between *U6* promters and gRNA scaffold for easy gRNA target site insertion. Target sites were identified using the web-based ZiFiT Targeter program (Hwang *et al.* 2013) or DRSC CRISPR (Ren *et al.* 2013). The following Target sites were used to target *yellow*, *Her*, *HLHm5*, *msh*, and *runt*:

y#1 GCGATATAGTTGGAGCCAGC;
y#2 GTGCACTGTTCCAGGACAAA;
Her#1 GCCGTTGTGTTGCAGAAATT;
Her#2 GGTGGTGAACTGCCAATTCC;
HLHm5#1 GCCATTCTTGAAGCTATCCA;
HLHm5#2 GGGATACCACAGCGATAACG;
msh#1 GGGATAAGTGGCGGCCCAGT;
runt#1 GGGGATCAGATGCCCTAGTA.

UAS-GFP::Dbox was created by a three fragment ligation of pUASt-attB (*Eco*RI/*Xba*I), GFP (*Eco*RI/*Xho*I, with pUASt-mCD8::GFP as its template)(Lee and Luo 1999), and *Drosophila* securin N50 a.a., covering the KEN-box and D-box (*Xho*I/*Xba*I; self-amplification from two primers) (Leismann and Lehner 2003).

Fly Strains

Fly strains used in this study were: (1) *bamP-GAL4::VP16* (gift of Dennis M Mckearin); *bamP-Gal4* in *attP2* (gift of Jon-Michael Knapp); (2) *10XUAS-IVS-GFP-p10* in *attP2* (pJFRC28) (Pfeiffer *et al.* 2012); (3) *nos.UTR-GAL4::VP16* (Bloomington Drosophila Stock Center, BDSC, #4937); (4) *UASt-FLP* in *su(HW)attP8*; (5) *UASp-FLP* in *su(HW)attP8*; (6) *Pfife* in *attP40*; (7) *BPfife* in *attP40, su(Hw)attP6, VK00002, su(Hw)attP5, attP2, VK00005, su(Hw)attP1, VK00027, VK00020*, and *VK00040*; (8) *act5C-Cas9* in *attP2A* (Port *et al.* 2014); (9) *dU6-gRNAs* against *yellow*, *Her*, and *HLHm5* in *attP2*; (10) *UAS-Cas9WT-p10* in *attP2*; (11) *UAS-CasD10A-p10* in *attP2*; (12) *GMR3-LexA::GADd* in *attP40* and *attP2*; (13) *nSyb-LexA::p65* in *attP16* and *VK00027*; (14) y^{1} w^{*}*/Dp(2;Y)G, P{hs-hid}Y; P{70FLP}23 P{70I-SceI}4A/TM3, P{hs-hid}14, Sb1* (BDSC, #25679) (15) y^{1} w^{*}*/Dp(2;Y)G, P{hs-hid}Y; P{70FLP}11 P{70I-SceI}2B* sna^{Sco}*/CyO, P{hs-hid}4* (BDSC, #25679) (16) *5X-ri6TS-Rac1*^{V12} and *5X-ri6TS-Rac1*^{V12}*(3xP3-RFP)* in *attP40* and *VK00027*; (17) *UAS-FLP-p10* in *attP18*; (18) *UAS-I-SceI-p10* in *su(Hw)attP8*; (19) *bamP-Cas9-2A-FLP-2A-I-SceI* in *su(Hw)attP8* and *attP2*; (20)

{donor} and *{donor, gRNA}* for all targetings in *attP40*; (21) *$y^{1}w^{67c23}$P{Crey}1b; sna^{Sco}/CyO* (BDSC, #766). (22) *UAS-GFP::Dbox* in *attP2*.

Fly genetics

To recover donor transgenic flies, we raised donor injected larvae (GENETIC SERVICES, INC. or Rainbow Transgenic Flies, Inc.) at room temperature, crossed the eclosed adults with *GMR3-LexA::GADd*, and searched for rough-eyed progeny as successful donor transformants.

Traditional Golic Heat Shock (Figure 2.5A): using *Her* targeting as an example, virgin {donor} females were first crossed to male *+/Y, hs-hid; 5X-ri6TS-$Rac1^{V12}$; hs-FLP, hs-I-SceI/TM3, Sb* (created by combining fly strains #14 and #16). Their eggs were collected in vials every 12 hours. For HS 1.5-3D scheme, vials were heat shocked at 37 °C for 1 hour, 1.5 days after the vial flipping (organisms aged 1.5-2Days), and again on day 2.5 (organisms aged 2.5-3Days). For HS 3.5-5D, vials were first heat shocked on day 3.5 (organisms aged 3.5-4Days) and then on day 4.5 (organism aged 4.5-5Days). The resulting founder females were crossed to *nSyb-LexA::p65* for lethality selection. Finally, eclosing candidates were subject to chromosome mapping and genomic PCR for confirmation.

Cystoblast Induction (Figure 2.5B): founder females were generated by crossing male *UAS-FLP-p10, UAS-I-SceI-p10/Y; 5X-ri6TS-$Rac1^{V12}$* to female *bamP-GAL4; {donor}*, and then collecting their virgin progeny with the correct

genotype. They were then crossed to *nSyb-LexA::p65* for lethality selection; the survivors were further subject to chromosome mapping and genomic PCR confirmation.

Cystoblast induction plus CRISPR/Cas9 & gRNA (Figure 2.5C): Additionally, two more transgenes (*UAS-Cas9* and *dU6-gRNA*, both in *attP2*) were added to the previous cystoblast induction setup. Crosses with *nSyb-LexA::p65* were performed for lethality selection, and surviving candidates further went through mapping and PCR confirmation.

Golic+ Crosses (Figure 2.5D): founder females were generated by crossing *{donor, gRNA}* to *5X-riTS-Rac1^{V12}* in *attP40*; *bamP198-CFI* (or *bamP898-CFI*) in *attP2* or *5X-riTS-Rac1^{V12}* in *attP40*; *bamP-Gal4* in attP2, *UAS-CFI* in *VK00027*. Their virgin progeny were then crossed to *nSyb-LexA::p65* for lethality selection. The survivors were further subject to chromosome mapping and genomic PCR confirmation.

Targeting designs and molecular characterization of target loci

Homologous arms of about 2 - 3 kb each were designed to completely knock-out *Her* and *HLHm5* and knock-in *msh-T2A-Gal4* and *runt-T2A-Gal4* (Figure 2.6A,B and 2.7A,C). The following primers were used for amplifying these arms and cloning them into pTL1 or pTL2.

Her_55NgoMIV: TACGGCCGGCTTATAAAATGGGCTTTTATTTCTTTAGTG;
Her_53SacI: TACGGAGCTCGACCATACAATCCCTGTATGC;
Her_35BamHI: CGGGATCCGCACAAGTAAAATACAAATGGGGAC;
Her_33MluI: TACGACGCGTCCTCTCGCTATCTCCCTTCG;

HLHm5_55AgeI: TACGACCGGTGTTTTGTAGCGTTCCTCATGTTGC;
HLHm5_53SacI: TACGGAGCTCGTCCCTTCACCATCATCGCTG;
HLHm5_35PmeI: TACGGTTTAAACGGTCCAAGTAGGAGGTCCTCTG;
HLHm5_33MluI: TACGACGCGTCTGATGGCACACAATTAGCAGGG.
msh_55NgoMIV: TAATTGCCAGCAATTTGCACCG;
msh_53AgeI: TACGACCGGTTCCCAGGTGCATCAGGC;
msh_35BamHI: CGGGATCCTAAGTGGCGGCCCAGTTG;
msh_33PmeI: AACAAATGCCCGCAATCAGCG;
runt_55*Age*I: ACGTACCGGTAAGTGACCCCCGATAAAGTGAAGTGCATACCGAG;
runt_53*Stu*I: ACGTAGGCCTGTAGGGCCGCCACACGGTCTTCTGC;
runt_35*Bam*HI: ACGTGGATCCTAGGGCATCTGATCCCCAAAAATCTGGAGGAATGAAG;
runt_33*Mlu*I: ACGTACGCGTTCTCAACCGCTTGTAGTCACCATTTAAGTTTTGGAC.

The following BAC clones were used: BACR48L05 for *Her*; BACR13F13 for *HLHm5*; BACR10L12 for *msh*; BACR50G05 for *runt*. T2A-Gal4 was generated by cloning Gal4 from pBPGAL4.2Uw-2 (Pfeiffer *et al.* 2010) into pTL2 (*Stu*I/*Sac*I) with a forward primer containing T2A coding sequence. The following primers were used in genomic PCR for *Her* and *HLHm5* knock-out and *msh-T2A-Gal4* and *runt-T2A-Gal4* knock-in candidate confirmation:

Her_g5: GTGATTGAGATATAGGCATACAGGG;
Her_m5: CATTCAGTCGCTTAAGATCCGAAG;
Her_m3: AATGTGCAGTTTTATTTCATGCCTC;
HLHm5_m5: CCCATGGATAGCTTCAAGAATGG;
HLHm5_m3: GGTGGAAGACAGGATTCAATGTC;
lexAop: GCAGTCGAGGTAAGATTAGACTAG;
HLHm5_g5: GTGGCTGAATGAGACTGGTGTCGAC;
Gal4-1: CACACGCTTGTTCAATACTACTCAG;
Gal4-2: GATACTCCACCGAACCCAAAGAAG;
msh_g1: CATCCACTGCATCCAATCCTAGTG;
msh_g2: GGCGTTAATATCAAGCTGTGATTTCG;
runt_g1: AATGGTGGTTGCTCGATATACCGATATATAC;
runt_g2: CGGATTCGGATTGGACGAGTTAAATTC.

Immunostaining and fluorescence microscopy

Ovaries were dissected in 1X Grace's Insect Medium, supplemented (GIBCO, Life Technology) and fixed for 10 min in 4% paraformaldehyde fixation solution. Fixation solution was prepared by mixing 20% w/v paraformaldehyde solution (Electron Microscopy Sciences) 1:4 to 1X PBS solution. After fixation, ovaries were first rinsed 3 times, then washed 3 times (5, 15, 30 min) in PBST (0.2% Triton X-100 in 1X PBS), and finally incubated with primary antibodies (diluted in PBST + 5% normal goat serum) overnight at 4℃. Afterwards, samples were further incubated at room temperature for 1 hr before rinsed 3 times and washed 3 times (20 min each) in PBST. Ovaries were then incubated with Secondary antibodies (diluted in PBST + 5% normal goat serum) for 3 hr at room temperature, rinsed 3 times and washed 3 times in an hour. For mounting, ovaries were transferred on glass slides, separated into ovariole strings, rinsed 1 time with 1X PBS, rinsed again in SlowFade Gold antifade reagent (invitrogen) and soaked in SlowFade for fluorescence imaging.

Washed, collected embryos were dechorionated in 50% bleach for 3 min. Then, they were thoroughly rinsed with water, dried, and transferred into 4% paraformaldehyde fixative (1 ml) plus equal volume of heptane for 30 min fixation on a shaker. Afterwards, the bottom aqueous phase was replaced with 1 ml of methanol, and embryos were devitellinized on a shaker for 1 min. The top heptane phase was aspirated away followed by 3 rinses of the bottom phase embryos in gradually diluted methanol/PBS solution (3:1, 1:1, and 1:3; v/v). These embryos went through an additional round of 4% paraformaldehyde

fixation. Next, they were rinsed 3 times, further washed in PBST for 30 min, and finally incubated with anti-GFP primary antibody overnight at 4°C. The next day, they were rinsed and washed through the same procedures, and incubated with secondary bodies overnight at 4°C before being rinsed, washed, stained with Hoechst 33342 for 10 min and mounted in SlowFade. Fluorescent signals of ovarioles and embryos were collected by confocal serial scanning using Carl Zeiss LSM710 microscope. Images were processed with Fiji then rotated and cropped with Keynote.

The following primary antibodies were used in this study: rat monoclonal anti-GFP, 1:100 (MBL International Corporation, D153-3) in Figure 2.2A; rat monoclonal anti-GFP, 1:1000 (NACALAI TESQUE, INC. 04404-84) in Figure 2.2C,D; rabbit polyclonal anti-GFP, 1:1500 (Molecular Probes, A11122) in Figure 2.7B,D, different anti-GFP antibodies were used because of availability; mouse monoclonal anti-Fasciclin III, 1:50 (Developmental Studies Hybridoma Bank, DSHB, antibody 7G10); rabbit polyclonal anti-DsRed, 1:1000 (Clontech, Living Colors, 632496); mouse anti-α-spectrin, 1:25 (DSHB, antibody 3A9). Secondary antibodies from Molecular Probes were used in 1:200 dilution: Alexa Fluor 488 goat anti-rat IgG (H+L); Alexa Fluor 568 goat anti-rabbit IgG (H+L); Alexa Fluor 647 goat anti-mouse IgG (H+L). For embryos, Cy3-conjugated goat anti-rabbit (1:200; Jackson ImmunoResearch, 111-165-144) and Hoechst 33342 (1:1000; InVitrogen) were used.

Table 2.1. Golic / Cystoblast induction / Cystoblast induction + CRISPR/Cas

Targeting Method		*{Her KO} in attP40*				*{HLHm5 KO} in attP40*			
		# of FF[a]	Correct GT	Non-Specific Insertion	Escapers	# of FF	Correct GT	Non-Specific Insertion	Escapers
HS1.5-3D		115	5(3)[b]	0	1(1)	150	9(2)	0	0
HS3.5-5D		95	0	0	2(2)	700	10(7)	7(2)	14(11)
bamP-Gal4 + UAS-FLP&I-SceI		180	0	17(13)	38(24)	—	—	—	—
bam-Gal4 + UAS-FLP&I-SceI&Cas9WT	*dU6-3-gRNA-Her(HLHm5)#1*	94	14(12)	15(15)	24(19)	90	66(47)	10(8)	21(17)
	dU6-3-gRNA-He(HLHm5)r#2	98	4(4)	11(11)	18(14)	90	59(45)	13(8)	24(23)
	dU6-2-gRNA-Her(HLHm5)#1	32	0	3(3)	9(9)	—	—	—	—
	dU6-2-gRNA-Her(HLHm5)#2	40	0	2(2)	6(6)	—	—	—	—
bam-Gal4 + UAS-FLP&I-SceI&Cas9^{D10A}	*dU6-3-gRNA-Her(HLHm5)#1*	90	1(1)	12(11)	18(17)	90	2(2)	11(11)	19(18)
	dU6-3-gRNA-Her(HLHm5)#2	92	0	23(21)	22(20)	90	4(4)	15(14)	16(15)
	dU6-2-gRNA-Her(HLHm5)#1	30	0	1(1)	5(5)	—	—	—	—
	dU6-2-gRNA-Her(HLHm5)#2	59	0	11(10)	9(9)	—	—	—	—

Various strategies were utilized to delete Her or HLHm5 with the same corresponding transgenic {donor}. Since one FF could sometimes yield multiple candidates, the number of candidates may exceed the number of candidate-producing FF.

[a]: FF: Founder Females;

[b]: # of candidate (# of candidate-producing FF) in the specific category

Table 2.2 Targeting *msh* and *runt* with different versions of Golic+

	{msh-T2A-Gal4 KI, gRNA} in attP40				*{runt-T2A-Gal4 KI, gRNA} in attP40*			
Targeting Method	# of FF	Correct GT	Non-Specific Insertion	Escapers	# of FF	Correct GT	Non-Specific Insertion	Escapers
bamP198-CFI	95	47(36)[a]	8(8)	15(14)	50	12	3	20
bamP898-CFI	30	37(18)	53(25)	5(4)	—	—	—	—
bamP-Gal4 + UAS-CFI	90	82(53)	27(22)	45(33)	—	—	—	—

Golic+ was used to knock-in T2A-Gal4 in both *msh* and *runt*. The first available Golic+ set with transgenes on II and III chromosomes was used although not ideal for targeting *msh* on III chromosome.

[a]: # of candidates (# of candidate-producing FF) in the specific category

Table 2.3. List of transgenic lines required for implementing Golic+

Full Name	Abbreviation	Integration Site	Note
donor DNA plus gRNA in pTL	{donor, gRNA}	attP40, VK00027	
GMR3-LexA::GADd	GMR3-LexA	attP40, attP2	Cross with {donor, gRNA} injected adults to create rough eyes for {donor} transformant screening.
bamP198-Cas9-P2A-FLP-E2A-I-SceI; bamP898-Cas9-P2A-FLP-E2A-I-SceI	bamP198-CFI; bamP898-CFI	su(Hw)attP8, attP2	Expressing Cas9, FLP and I-SceI under the bamP control to release donor DNA and introduce DSB at the target locus in every cystoblast.
5XLexAop2-rCD2miRNATS#6-Rac1^{V12} (3xP3-RFP)	5X-riTS-Rac1^{V12}	attP40, VK00027	Together with {donor, gRNA}*, providing a homozygous suppressible “Toxic” background.
Residual {donor, gRNA}	{donor, gRNA}*		After donor release, it will reconstitute as a suppressible toxic module, 5X-FRT-riTS-Rac1^{V12}.
nSyb-LexA::p65	nSyb-LexA	attP16, VK00027	Larval/pupal lethality selection

Table 2.4. Quick reference for using Golic+

	Targeting		
Transgenes set	X Chromosome	2nd Chromosome	3rd Chromosome
{donor, gRNA} in	attP40	VK00027	attP40
5X-riTS-Rac1^{V12} in	attP40	VK00027	attP40
nSyb-LexA in	attP16	VK00027	attP16
bam198-CFI in	attP2	su(Hw)attP8	su(Hw)attP8

Figure 2.1. Comparison between different gene targeting strategies.

Ends-out GT steps: generation of linear donor DNA, homologous recombination between donor DNA and targeted sequences, and recovery of correct GT. (A) For the Golic heat-shock strategy, donor is first inserted in the genome as a P-element transgene, and then released (flip-out and linearization) in larval primordial germ cells by heat shock-induced expression of FLP and I-SceI. Targeting occurs rarely through endogenous homologous recombination machinery. Candidates, possibly carrying the same GT event due to later clonal expansion, are recovered based on the mini-white eye marker in-between the 5' and 3' homology arms. (B) For embryo microinjection, donor DNA is injected together with the corresponding sequence-specific nuclease to boost GT in embryonic pole cells. Clonal expansion can again lead to multiple offspring carrying identical GT. (C) In Golic+, donor DNA is not released from the transgene until the birth of each cystoblast (CB) from the ovarian germline stem cells, guaranteeing independent GT among candidates. Ends-out GT in CBs requires DNA double strand breaks made by CRISPR/Cas, and recovery of correct GT is facilitated by a repressor-based lethality selection. The CB-specific induction of FLP, I-Sce1, and Cas9 depends on *bamP-Gal4*; guide RNA for CRISPR/Cas is broadly expressed with the *dU6* promoter; strong candidates are recovered based on inheritance of a repressor, miRNA against rCD2, to rescue the pupal lethality caused by nSyb-driven riTS-Rac1^{V12}.

Figure 2.1. Comparison between different gene targeting strategies.

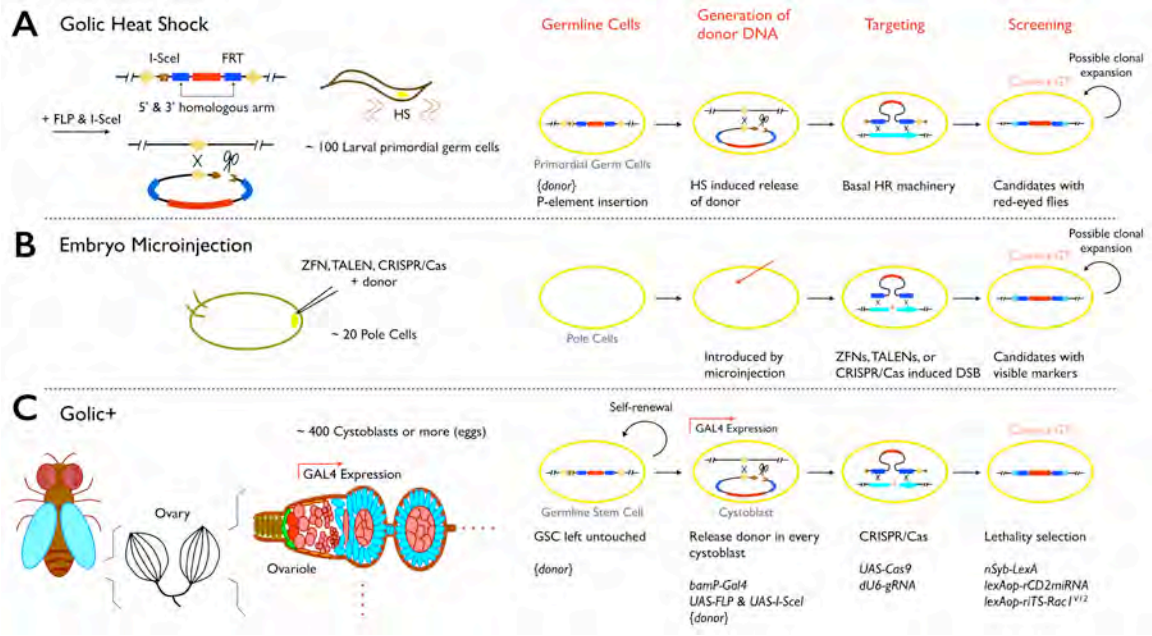
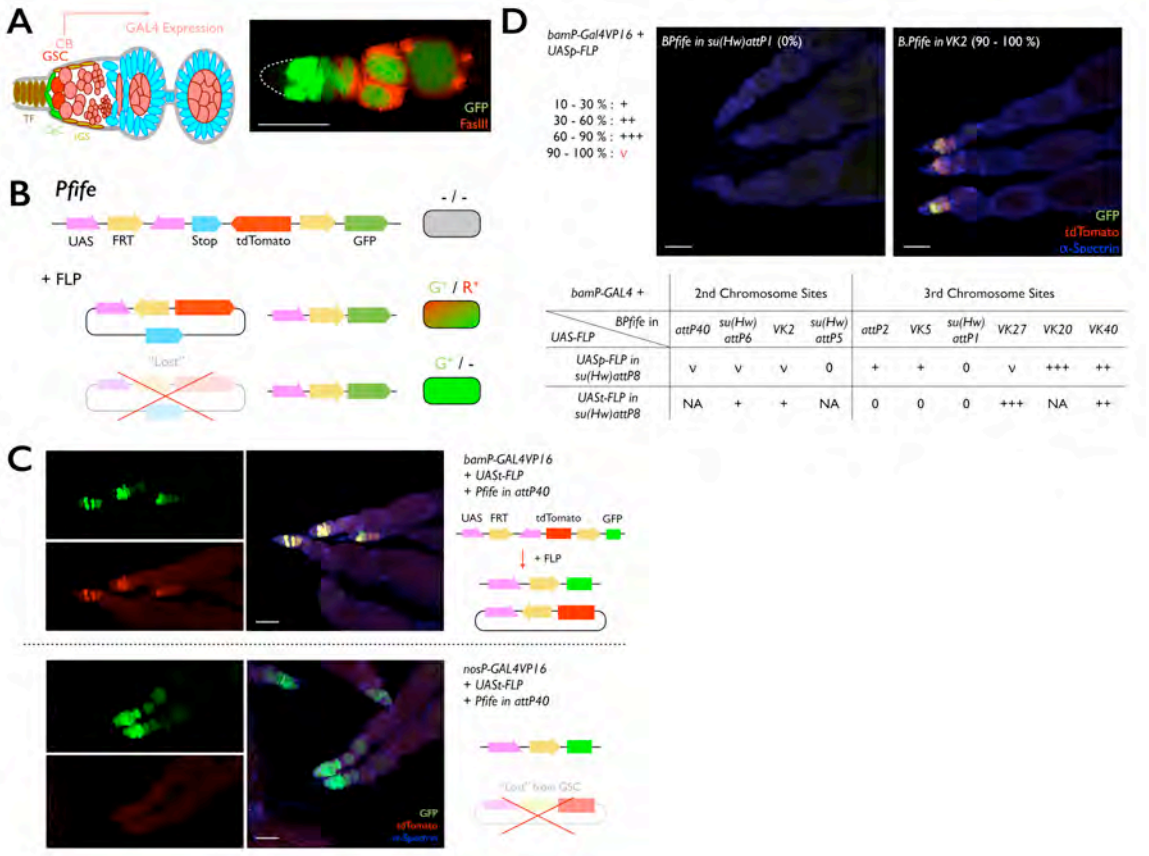


Figure 2.2. Optimizing cystoblast-specific excision of FRT cassettes.

(A) *bamP-GAL4::VP16* drives GFP expression in newborn CBs, but not in the preceding GSCs located at the tip of the germarium outlined with anti-FasIII immuostaining (red). TF: terminal filament; GSC: germline stem cell; CB: cystoblast; CpC: cap cells; IGC: inner germarium sheath cell (B) Pfife carries two interrupted UAS-reporters that are reconstituted based on FLP based recombination at the FRT sites. Upon flip-out, UAS-GFP is expressed by the residual Pfife and UAS-tdTomato is expressed by the circularized FRT cassette. Co-expression of GFP and tdTomato indicates persistence of the excised cassette following flip-out, while GFP alone reports older flip-out events where the circularized FRT cassette has been lost through cell division. (C) Induction of flip-out in ovarioles using *bamP-GAL4::VP16* versus *nosP-GAL4::VP16*. *bamP* elicited flip-out occurs specifically in cystoblasts that were persistently labeled with GFP plus tdTomato, while the *nosP*-mediated flip-out occurred in GSCs prior to adult oogenesis as indicated by expression of GFP alone throughout the female germline. (D) The larger BPfife placed at various attP sites on second and third chromosomes was assayed for flip-out mediated by *bamP-GAL4::VP16*-driven *UASp*- versus *UASt-FLP*. As summarized in the table below, flip-out efficiency varied drastically with the insertion site, and *UASp-FLP* outperformed *UASt-FLP*. Note: 0% flip-out in *su(Hw)attP1* versus almost 100% in *VK2* on a per ovariole basis . Scale bars: 50 μm.

Figure 2.2. Optimizing cystoblast-specific excision of FRT cassettes.



bamP-GAL4 + UAS-FLP \ BPfife in	2nd Chromosome Sites				3rd Chromosome Sites					
	attP40	su(Hw) attP6	VK2	su(Hw) attP5	attP2	VK5	su(Hw) attP1	VK27	VK20	VK40
UASp-FLP in su(Hw)attP8	v	v	v	0	+	+	0	v	+++	++
UASt-FLP in su(Hw)attP8	NA	+	+	NA	0	0	0	+++	NA	++

Figure 2.3. Transgenic CRISPR/Cas system.

(A) Sequence arrangements of the dU6-gRNA scaffolds are shown on top. dU6-gRNA backbone was cloned into pJFRC28 using *Hin*dIII and *Eco*RI sites. Two *Sap*I sites were put in-between *dU6* promoter and the gRNA scaffold for easy target site cloning. gRNA scaffold is the same as the published one (Cong *et al*. 2013; Mali *et al*. 2013). After annealing two corresponding target site primers, the target site can be directly ligated with the SapI-digested empty dU6-gRNA with its TCG and AAG 5' overhangs to constitute a functional dU6-gRNA. (B) Females carrying various *U6-gRNAs* against *yellow* were crossed to males with *act5C>Cas9*. Their female progeny showed allele-dependent yellow body color mosaicism, with *dU6-3-gRNA-y#1* causing a yellow phenotype throughout almost the whole body whereas *dU6-2-gRNA-y#2* affected few, if any, cells. (C) Summary of yellow mutations in the female germline using *bamP-GAL4* driven Cas9 or $Cas9^{D10A}$, plus four kinds of *dU6-gRNAs* against *yellow*. Two *yellow* mutants show small deletions around the cleavage sites of target site #1 and #2.

Figure 2.3. Transgenic CRISPR/Cas system.

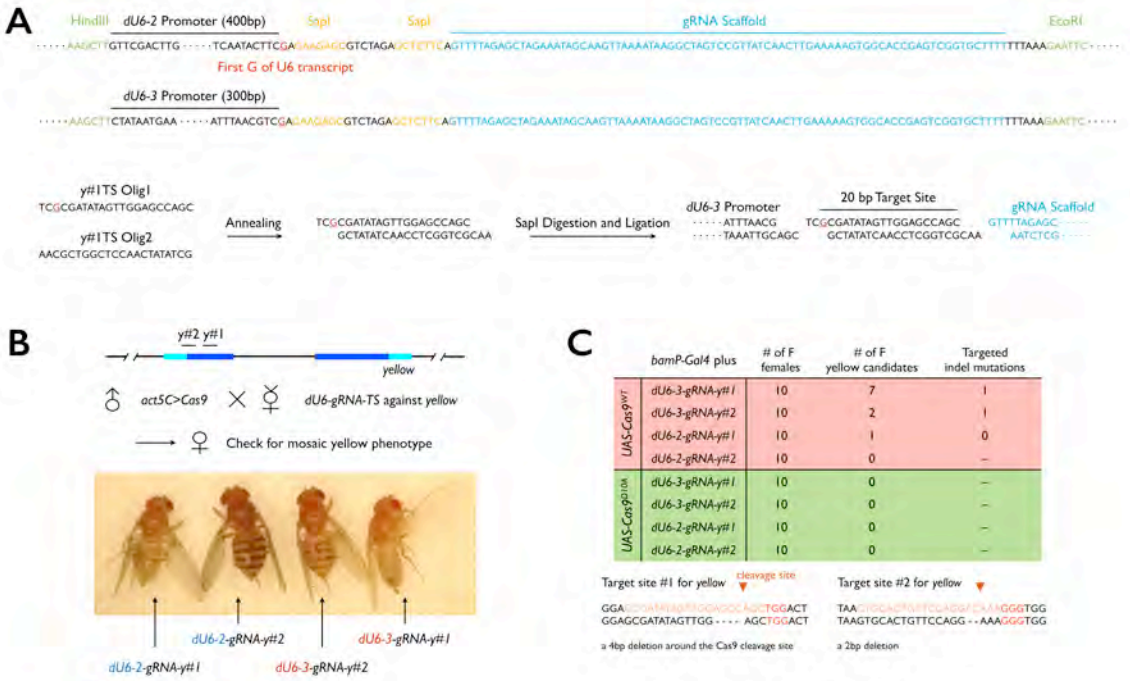


Figure 2.4. Repressor-based selection and GT plasmids.

(A) Rough to bar eye and dead pupa phenotypes were elicited by *GMR>Rac1*^{V12} and *nSyb>Rac1*^{V12}, respectively. (B) pTL1 has a similar organization to pRK1. Note the presence of lexAop-rCD2i, instead of hsp70::white, as the marker residing between the 5' and 3' multiple cloning sites (MCS), and use of LexAop2-opRac1^{V12}, rather than UAS::Rpr, for eliminating non-specific insertions. Two more modules, LexAop2 and riTS-Rac1^{V12}, are separated by the FRT cassette for reconstitution of a suppressible LexAop2-riTS-Rac1^{V12} following flip-out. pTL2 is pTL1 plus a *dU6-3* promoter driven gRNA. Two SapI sites were included for easy gRNA target site cloning. In pTL2, a second I-SceI site, which is unnecessary for linearization and may have increased the frequency of incorrect excisions events, was also removed. (C) {donor} is first integrated at attP sites on either second or third chromosome. To enrich for correct GT following induction of donor DNA flip-out, three scenarios of "unwanted" events are eliminated by lethality selection. Non-suppressible Rac1^{V12} is expressed under *syb-LexA::p65* to kill organisms experiencing "no excision" or "non-specific insertion". In the case of "no targeting", the reconstituted LexAop2-riTS-Rac1^{V12} drives lethality. By design, only after ends-out GT leading to loss of the non-suppressible LexAop2-opRac1^{V12} can the organism overcome the suppressible LexAop2-riTS-Rac1^{V12} with lexAop-rCD2i and survive the selection.

Figure 2.4. Repressor-based selection and GT plasmids.

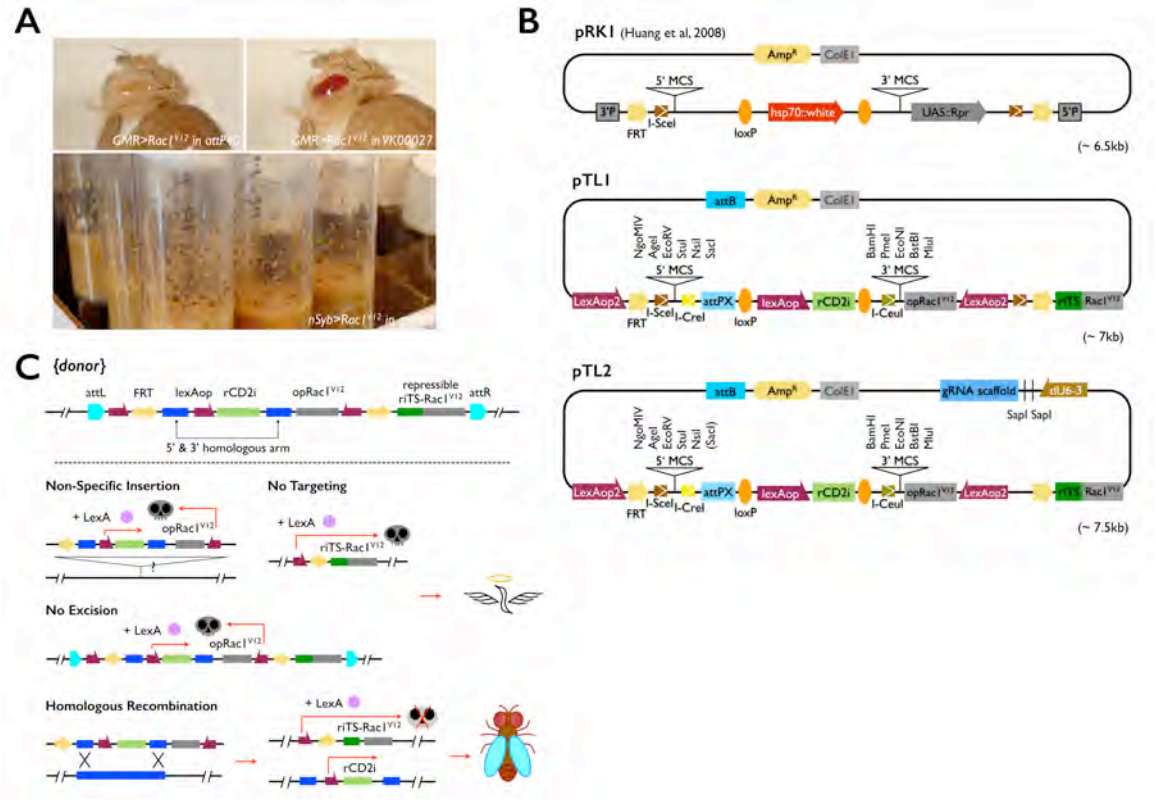


Figure 2.5. Crossing schemes for different GT strategies.

All targeting strategies in this study consist of two major crosses and three steps (donor release & targeting, lethality selection, and chromosomal mapping and genomic PCR). The first cross is to generate linear {donor} with FLP and I-SceI (and additionally bring in Cas9 and gRNA), and hence create the founder females. The second cross is to perform the lethality selection by mating founder females to male *nSyb-LexA::p65*. Eclosed candidates are collected for genetic mapping and PCR confirmation. (A, B, C) In these three schemes, real targeting at *Her* is denoted as *Her**. Gene targeting for *HLHm5* underwent the same genetic crosses, only that *HLHm5** would have been indicated at the third chromosome. (D) Knocking T2A-Gal4 into *msh* with Golic+. The first available Golic+ set with transgenes on II and III chromosomes was used although not ideal for targeting *msh* on III chromosome.

Figure 2.5. Crossing schemes for different GT strategies.

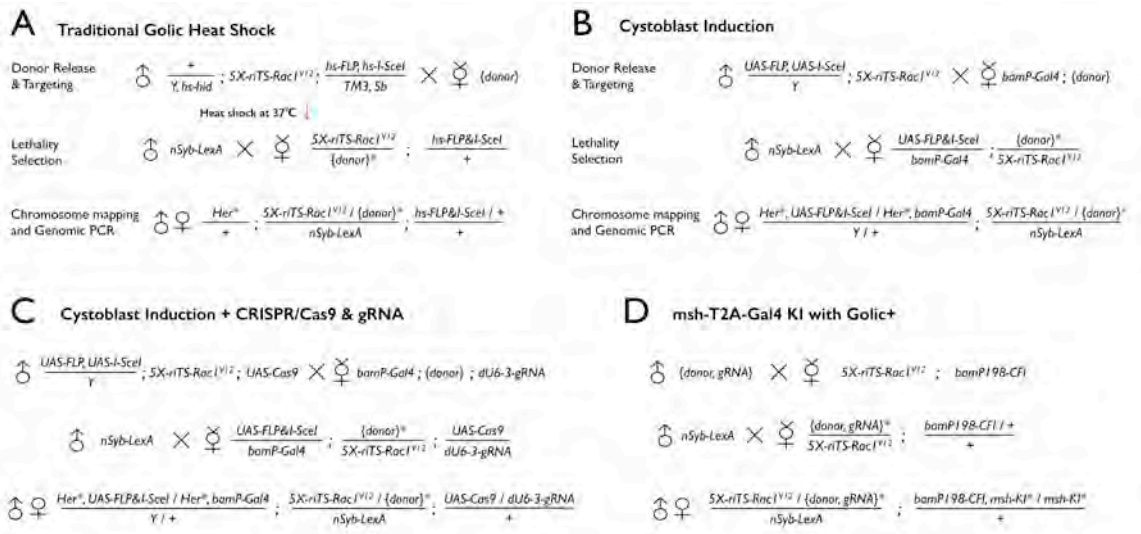


Figure 2.6. Characterization of GT candidates.

(A, B). Homologous arms (~3kb each) were cloned into pTL1 to knock out *Her* and *HLHm5*. *Frq2* is a gene adjacent to *Her*. Target sites against *Her* and *HLHm5* are indicated as short lines above the genes. To confirm *Her* knock-outs, three primers (Her_g5, Her_m5, and Her_m3) were used for genomic PCR. With successful targeting, these primer pairs (g5m3, m5m3) should yield no PCR fragments. The same idea applied to the HLHm5_m5 and HLHm5_m3 primer pair. Meanwhile, we design two primers (lexAop, located within the lexAop-rCD2i suppressor; HLHm5_g5, located upstream of the 5'arm for *HLHm5*) to examine the *HLHm5* locus before removing the suppressor *loxP* cassette. Amplification of the HLHm5 g5/loxAop PCR fragment (~3.9kb) is only possible when the suppressor is integrated in the *HLHm5* locus through HR. (C) Local integrations refer to those retaining the repressor yet losing the non-repressible toxic module and hence surviving the lethality selection, possibly due to local hopping given their presence on the chromosome where {donor} originates from. By contrast, escapers have eclosed without the repressor-marked GT DNA due to failure in the reconstitution of a functional repressible toxic module at the {donor} residual site, apparently because of imprecise flip-out or premature I-SceI cutting.

Figure 2.6. Characterization of GT candidates.

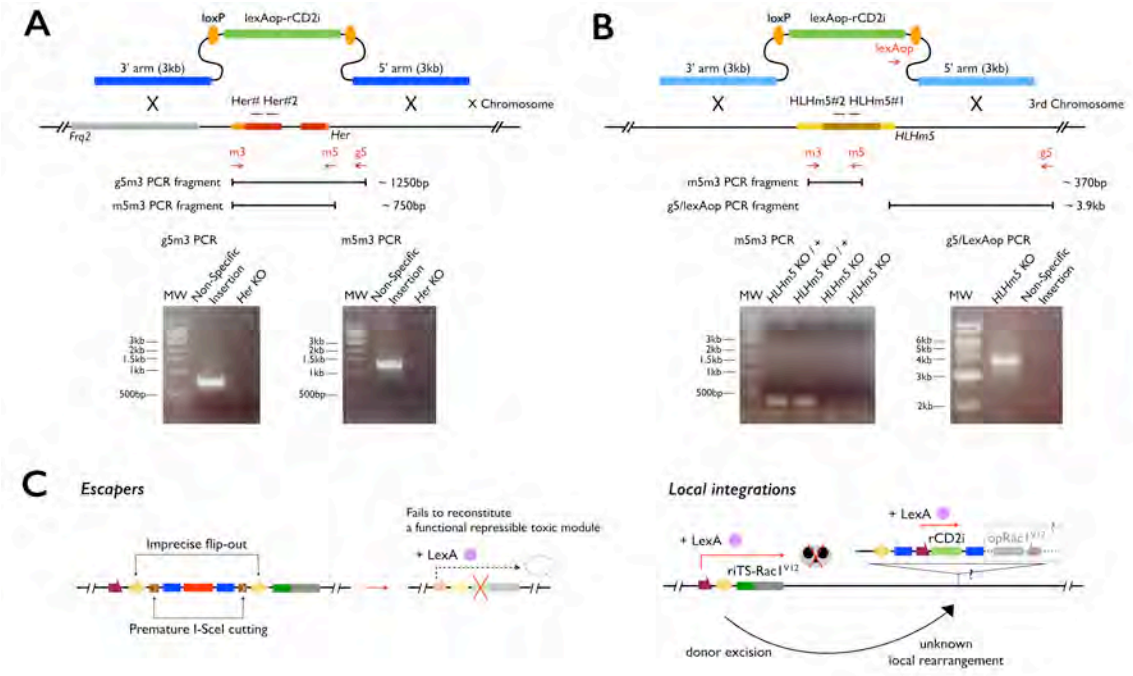


Figure 2.7. Targeting *msh* and *runt* with Golic+.

(A, C) Donor designs and PCR confirmation for generating *msh-T2A-Gal4* and *runt-T2A-Gal4* knock-ins. Genomic sequences (2-3kb) right upstream and downstream of the *msh* or *runt* stop codons were used as homology arms (TAA was included in the 3'arm). Coding sequence of T2A-Gal4 was placed in between so that GAL4 and Msh (or GAL4 and Runt) can be translated from the same mRNA transcript. gRNA target sites sitting around the stop codons were selected. For each gene, two sets of primers (msh_g1 and Gal4-1; msh_g2 and Gal4-2; runt_g1 and Gal4-1; runt_g2 and Gal4-2) were chosen to confirm T2A-Gal4 knock-ins at both ends. g1 and g2 primers were chosen further upstream or downstream of the homology arms so that PCR amplicons (3.1, 3.2kb for *msh*; 2.5, 2.4 for *runt*) were only possible when T2A-Gal4 is correctly situated at the *msh* or *runt* locus. The g2/Gal4-2 PCRs were performed after Cre removed the lexAop-rCD2i loxP cassette. (B, D) Bi-lateral patterned GFP expression seen in embryos of *msh-T2A-Gal4* and *runt-T2A-Gal4* knock-ins matches *msh* and *runt* BDGP in-situ data (http://insitu.fruitfly.org/cgi-bin/ex/report.pl?ftype=1&ftext=CG1897 and http://insitu.fruitfly.org/cgi-bin/ex/report.pl?ftype=1&ftext=CG1849) (Hammonds *et al.* 2013). Nuclear DNA was revealed by Hoechst 33342 staining. Scale bars: 50 μm. In all panels illustrations are not to scale.

Figure 2.7. Targeting *msh* and *runt* with Golic+.

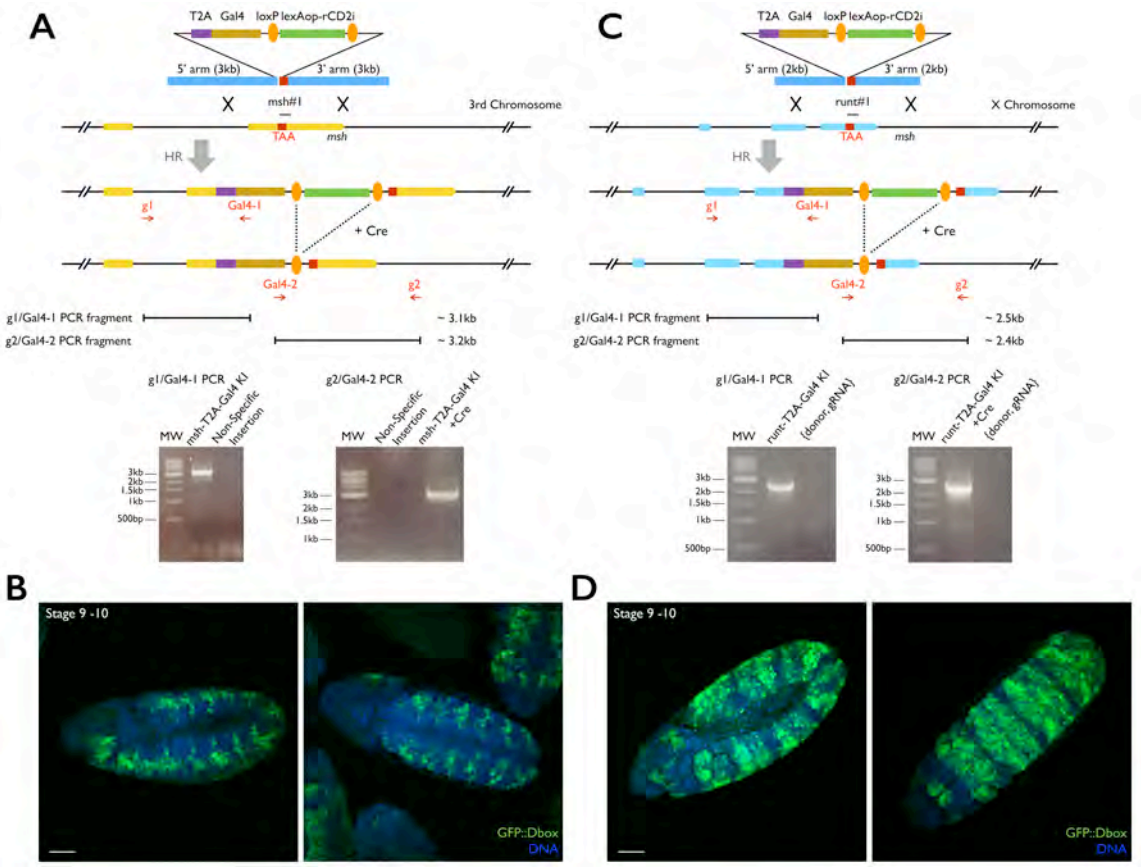


Figure 2.8. Golic+.

Golic+ founder females can be raised en masse in bottles because the linear donor DNA is released in each cystoblast for independent GT trials. Ends-out GT is greatly boosted by CRISPR/Cas and candidates are selectively recovered via suppression of *nSyb-LexA::p65*-mediated pupal lethality.

Figure 2.8. Golic+.

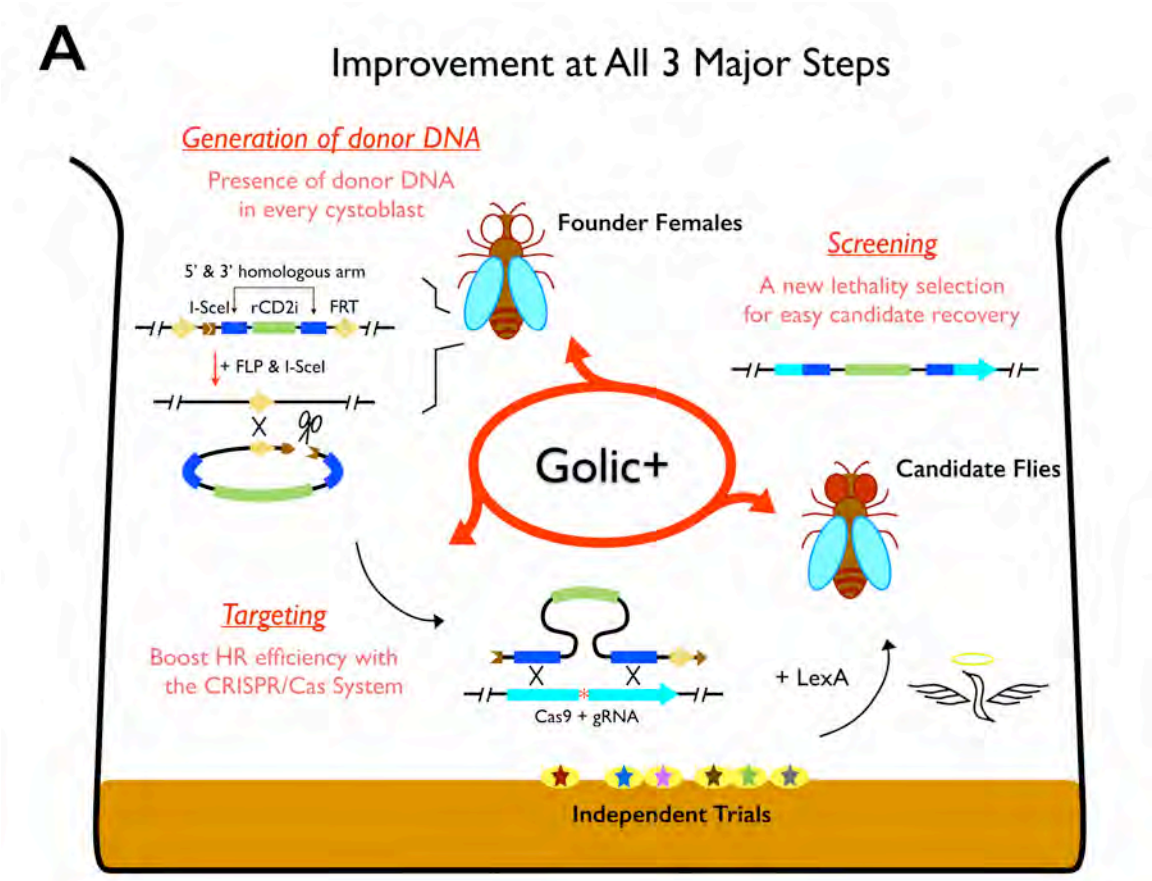
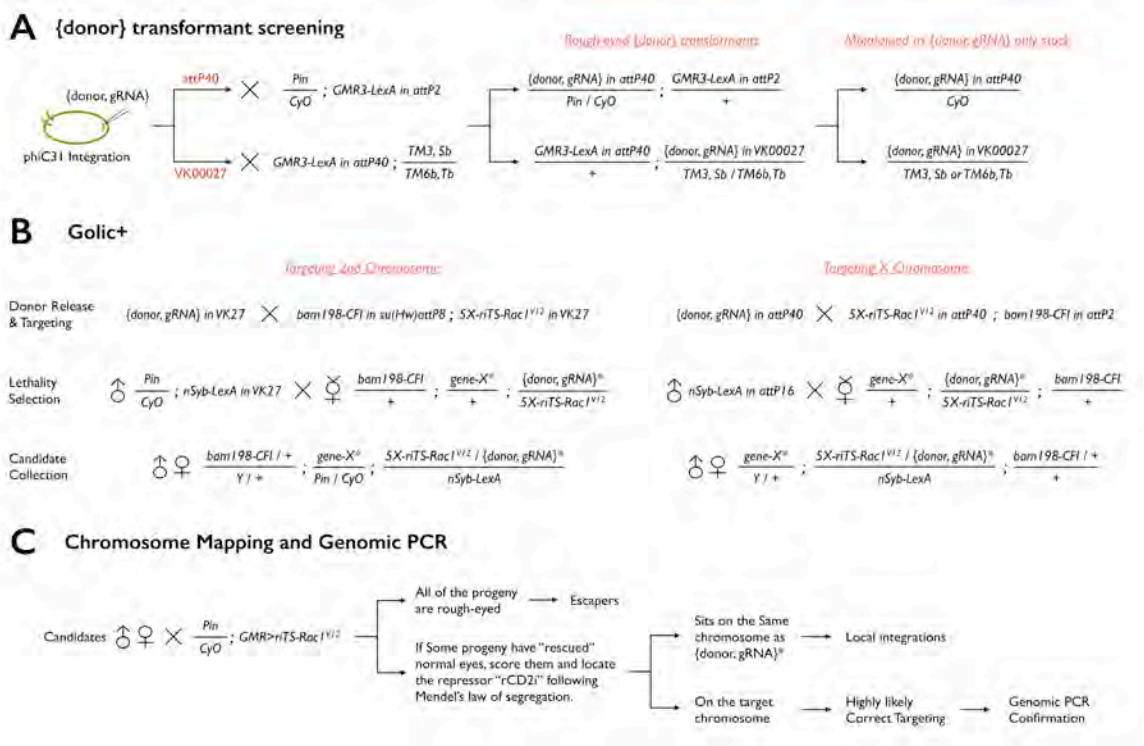


Figure 2.9. Crossing schemes for applying Golic+.

(A) {donor, gRNA} injected embryos were raised and then mated with *GMR3-LexA.* Transformants with obvious rough bar eye phenotype were identified and maintained as balanced stocks. (B) Schemes for targeting a 2nd or a X chromosome gene are depicted here. Targeting a 3rd chromosome gene can be deduced from referencing Table 2. Golic+ consists of two major crosses and three steps (donor release & targeting, lethality selection, and candidate collection). The first cross is to generate linear {donor} with FLP and I-SceI (and additionally express Cas9 and gRNA), and hence create the founder females. The second cross is to perform the lethality selection by mating founder females to male *nSyb-LexA::p65*. Eclosed candidates are collected for genetic mapping and PCR confirmation. (C) A flowchart to characterize GT candidates. Crossing candidates to GMR>riTS-Rac1^{V12} allows deduction of the presence and location of the rCD2i repressor. Candidates with the repressor relocating to the target chromosome likely carry correct GTs. Those with the repressor remaining on the original chromosome are categorized as local integrations. Escapers, by contrast, carry a defective 5XLexAop2-FRT-riTS-Rac1^{V12} and have therefore escaped the lethality selection. The escapers without the rCD2i repressor show no suppression of the rough-eyed phenotype in their progeny.

Figure 2.9. Crossing schemes for applying Golic+.



CHAPTER III

Tracing specific NB lineages based on early patterning genes.

1) This chapter is unpublished. Likely authors: Hui-Min Chen, Ching-Po Yang, Haluk Lacin, Josephine Chang, and Tzumin Lee. Hui-Min Chen wrote this chapter.
2) Tzumin Lee conceived the project. Hui-Min Chen designed and generated the T2A-Gal4 knock-ins. Josephine Chang made the donor constructs. Ching-Po Yang characterized T2A-Gal4 knock-in expression at the embryonic, larval stages. Haluk Lacin characterized T2A-Gal4 knock-in expression in embryonic ventral neurogenic region. Ching-Po Yang performed the immortalization experiments, and Hui-Min Chen mapped the hit lineages by stochastic clonal labeling.
3) Embryonic in situ images of *lab*, *unpg*, *hkb*, *vnd*, *ind*, *msh* were modified from BDGP insitu database (http://insitu.fruitfly.org)

Introduction

The *Drosophila* CNS develops from a strictly patterned two-dimensional sheet of cells, the neuroectoderm, into a massively integrated three-dimensional structure consisting of the supraesophageal ganglion (SPG), the subesophageal ganglion (SEG), and the truncal ventral nerve cord (VNC). The cerebral part of SPG can be divided into three segmental neuromeres: protocerebrum, deutocerebrum, and tritocrebrum, which come from the pregnathal segments: labral/ocular, antennal, and intercalary respectively. As for the VNC, it is made up of 3 gnathal (mandibular, maxillary, and labial), 3 thoracic, and 8 abdominal neuromeres (Schmidt-Ott *et al.* 1994).

These CNS neuromeres are composed of numerous diverse and distinct neurons deposited by around 1000 NBs from embryonic to pupal stages. Each fly body segment can be split into two left-right mirrored hemisegments that generally contain 30 NBs each. However, in the anterior cerebral segments, NB numbers deviate from the common standard drastically. The most anterior labral neuromere comprises about 10 NBs, while the much larger ocular neuromere has more than 60 NBs. The antennal and intercalary neuromeres contain 21 and 13 NBs, respectively (Urbach *et al.* 2003).

Amazingly, owing to the repetitive nature of the insect body plan, NBs positioned at corresponding spots in different VNC segments, termed serially homologous NBs, have relatively similar identities (Doe 1992). Within each

classic VNC segment, anterior/posterior (AP) and dorsal/ventral (DV) patterning genes are differentially expressed in the neuroectoderm. On the AP axis, cell fates are continually defined and resolved by the classic cascade actions of gap genes, pair-rule genes, and finally segment polarity genes, such as *wingless*, *hedgehog*, *patched*, *gooseberry*, *engrailed*, *mirror*, and *invected* (Akam 1987; Skeath and Thor 2003; Technau et al. 2006). As a result, every segment can be distinctly subdivided into four parallel rows by different combinations of segment polarity genes (Bhat 1999; Skeath 1999; Technau *et al.* 2006). On the other hand, along the DV dimension, three instructive signals, nuclear factor NF-κB, bone morphogenetic protein (BMP), and epidermal growth factor receptor (EGFR) signaling pathways, pattern the expression of columnar genes, including *ventral nervous system defective* (*vnd*), *intermediate neuroblasts defective* (*ind*), *muscle segment homeobox* (*msh*), and *Drosophila EGF receptor* (*DER*), which in turn divide the neuroectoderm into three longitudinal columns (Skeath 1998; Bhat 1999; Von Ohlen and Doe 2000; Skeath and Thor 2003; Technau *et al.* 2006). Together, segment polarity genes and columnar genes neatly establish a repetitive orthogonal Cartesian coordinating system in every VNC hemisegment. Nevertheless, each hemisegment does have its own characteristics. This comes from the effects of different homeotic genes (or called Hox genes) acting along the body length (Hiromi and Gehring 1987; Carroll 1995). The segment-specific expressions of Hox genes consequently further diversify the serially homologous

NBs and provide them each with a unique identity (Hirth *et al.* 1998; Rogulja-Ortmann and Technau 2008).

Lessons learned from the VNC studies helped the Technau group initiate their effort to survey expression of known important “molecular markers” in the pregnathal neuromeres (Urbach *et al.* 2003; Urbach and Technau 2003b; Urbach and Technau 2003a). However, the orderly AP/DV patterning principles are mostly lost in the abstract pregnathal segments, likely the result of an evolving process to support their higher computational functions (Bally-Cuif and Wassef 1995; Urbach and Technau 2004; Urbach 2007; Pani *et al.* 2012). Generally speaking, the orthogonal expression of segment polarity and dorsal/ventral patterning genes is roughly conserved in the posterior region, but it gradually becomes obscure towards the anterior end (Urbach and Technau 2003b). The tritocerebrum behaves like a “reduced” trunk neuromere; the deutocerebrum has a less conserved expression profile of patterning genes; the protocerebrum seems to abolish most of the rules and hence coordinates. Expectably, NBs within the pregnathal neuromeres do not arise in a readily comprehensible manner.

Meanwhile, the distinct and elaborate morphologies of the cerebral lineages are one of the responsive end results of the original intricate NB patterning (Pereanu and Hartenstein 2006; Yu *et al.* 2013a). While systematically grouping and cataloging cerebral lineages by their characteristic cell body locations and neurite trajectories, it is observed that neighboring lineages often

coinnervate the same neuropil or neuropils and project to restricted set of distant neuropils. One obvious example is the five antennal lobe (AL) lineages. Even though complex and distinguishable themselves, they all have dendritic elaborations in the AL region while extending their axonal trajectories to lateral horn (LH), mushroom body (MB) calyx, and other neuropils. Such phenomena immediately suggest the existence of common factors among lineages to direct shared morphological features, plus unique determinants to produce specialized morphologies.

Inspired by Technau group's effort, we embarked on our journey to study the hidden logic behind the *Drosophila* brain development. Though complicated at first glance, Urbach and Technau's NB "fate" map is extremely valuable as the fundamental framework for investigating tissue patterning and cell diversification in the *Drosophila* brain, and also for studying the differentiation of the brain from its truncal counterpart. We intend to utilize this map by linking the molecular markers' expression to the adult lineages that previously expressed them at the early embryonic stage when NBs delaminate. These links have previously been difficult to establish due to the lack of a simple and faithful strategy to continually follow the corresponding progenitors and their progeny. Thanks to the recent efforts by Awasaki *et al.*, we now have the capacity to convert drivers that transiently express in NBs into drivers that express in all of their subsequent progeny, a process termed immortalization (Awasaki *et al.* 2014). Briefly, immortalization involves intersection of the activities of a pan-NB specific *dpn*

promoter and a candidate driver transiently expressing in NB. This overlap in turn generates an irreversible recombination event in the progenitor and makes all of its progeny uniformly labeled or influenced, hence targeting specific lineages.

Equipped with the immortalization technology, we simultaneously developed our own efficient gene targeting (GT) strategy to generate a good collection of “molecular marker” drivers for specific lineage targeting. Additionally, the discovery of the 2A peptide modules, which promote ribosomal skipping during translation, makes it possible to piggyback a foreign coding sequence (CDS) to an endogenous gene and hijack its expression profile (Szymczak *et al.* 2004; Diao and White 2012). Combining these two technologies and immortalization together, we hope to create numerous T2A-Gal4 KIs and extrapolate Urbach and Technau’s map to a molecular marker/cerebral lineage correspondence table. Describing such a link is the first step to deeper appreciation of their cause-and-effect relationships. In addition, with more advanced genetic intersection, we believe targeting specific NB/lineage is achievable for informed mechanistic developmental studies.

Results

Generating T2A-Gal4 Knock-ins for genetic fate mapping

The mystery remains in how the *Drosophila* brain discards the seemingly logical orthogonal VNC blueprint and yet manages to assemble a much more sophisticated structure. We started with analyzing the developmental logic of the two relatively organized posterior pregnathal neuromeres, tritocerebrum and deutocerebrum to assess the practicability of our proposal. Six molecular markers (*lab*, *unpg*, *hkb*, *vnd*, *ind*, and *msh*) were accordingly selected for their enriched expression in tritocerebral and deutocerebral NBs (Urbach and Technau 2003a).

lab is expressed in all 13 tritocerebral NBs in Urbach and Technau's table. We are interested in knowing the adult identities of these tritocerebral lineages because they might be positioned along the SPG/SEG border, which has never been strictly defined before owing to the mingling of neuromeres during metamorphosis. Two other genes, *huckebein* (*hkb*), and *unplugged* (*unpg*) were selected for their concentrated expressions in NBs of the adjacent deutocerebrum. Lastly, *vnd*, *ind*, and *msh* were obvious choices because of their preserved differential DV distribution in both tritocerebrum and deutocerebrum.

Donors to knock in T2A-Gal4 in each of the endogenous genes were created with lengths of both of the homology arms set at about 3kb. Due to our initial lack of experience, long arms were chosen to improve the chance of

correct targeting. Five KIs (*unpg-*, *hkb-*, *lab-*, *ind-*, and *vnd-T2A-Gal4*) were generated before Golic+ was mature, and hence they are products of traditional heat shock strategy with pTL1 as their donor backbone. We successfully recovered correct targeting for all five KIs, and observed efficiencies of roughly 1 correct targeting per 100 founder females. Only *msh-T2A-Gal4* was created by Golic+ with an amazing recovery rate of 1/2 (correct targeting/founder female) due to the effect of CRISPR/Cas induced target-specific DSB (see Chapter II). The success of all 6 GT cases also indicates the robustness of the lethality selection design.

Correct targetings were immediately confirmed by chromosomal mapping and generation of correct genomic PCR amplicons. We used *UAS-GFP::Dbox* to evaluate their Gal4 activities and found they faithfully generated expression patterns that matched up with the *in situ* data from BDGP (Hammonds *et al.* 2013), with a beautiful ventral to dorsal progression of expression of the three columnar genes, *vnd*, *ind*, and *msh* (Figure 3.1). These molecular markers also continue to be selectively expressed in several NBs through larval stages (Figure 3.1), which suggests continual requirements of these genes in the development of certain lineages beside initial NB fate determination.

To further validate our KI drivers, we inspected expression produced by *vnd-*, *ind-*, and *msh-T2A-Gal4* in the well-studied embryonic vNR. As expected, *vnd-*, *ind-*, and *msh-T2A-Gal4* drive reporter expression in medial, intermediate, and lateral NB columns of stage 11 embryonic vNR, respectively (Figure 3.2A-C).

At earlier stage 9, when only three distinct columns of NBs appear, *vnd-* and *ind-T2A-Gal4* promote expression in medial and intermediate column NBs (Figure 3.2D,E). Surprisingly, *msh-T2A-Gal4* does not induce obvious expression in the lateral NB columns at this stage (data not shown). Lateral expression by *msh-T2A-Gal4* only becomes detectable later, and overlaps well with the Msh antibody staining in several NBs (Figure 3.2F).

Despite the usefulness of these KI drivers, we were surprised to observe that some of the knock-in Gal4s lost their activities during routine stock maintenance. Most of the T2A-Gal4 KIs do not behave as fully wild-type alleles from the fact that they are often maintained as heterozygotes over balancer chromosomes. We suspect the T2A module might not promote 100% ribosomal skipping, and the hybrid proteins possibly cause unexpected ectopic toxicity for the organisms. Because the expression of a T2A-Gal4 is contingent upon an intact continuous CDS starting from its target gene, its translation can be easily disrupted by small indels or missense mutations. Additionally, a truncated CDS is likely still more beneficial than an ectopic hybrid transcription factor, so the further mutated population gradually compete out the original knock-in population, though still appear heterozygous and balanced. Preliminary genomic sequencing data does support this hypothesis with the discoveries of small deletions in several T2A-Gal4 KIs that lost their binary induction abilities.

Lineage tracing by immortalizing NB expressions

First, we intended to describe the overall lineage coverage from each of our 6 T2A-Gal4 KIs. The following immortalization strategies were achieved owing to the discovery of a pan-NB specific *dpn* promoter that is active in every NB outside of the optic lobes and silent in GMC and post-mitotic neurons (Awasaki *et al.* 2014). To capture and maintain the early NB expressions of these KIs, we first relayed their Gal4 activity to UAS-FLP, which subsequently removes a stop cassette (>FRT-stop-FRT>) between a *dpn* promoter and a secondary driver, LexA::p65 (Figure 3.3A). Then, the reconstituted *dpn-FRT-lexA::p65* could turn on *LexAop2-myr::GFP* to label specific NBs that correspond to the original knock-in T2A-Gal4 activities. Noticeably, neurite tracts of adjacent progeny cells from these NBs were also visible due to the perdurance of reporter expression (Figure 3.3B-G).

We were surprised to see only 2 NBs that ever express *lab* (denoted *lab*+ NBs) under this genetic intersection (Figure 3.3B), which suggests the majority of tritocerebral NBs do not survive into wandering (WD) larval stage. Also, we found, instead of covering almost all of the deutocerebral NBs as originally reported (Urbach and Technau 2003a), there are only 2 *unpg*+ NBs immortalized (Figure 3.3C), which is in consistent with Urbach's following corrected annotation (NB Dv7 and Dd1) (Urbach 2007). *hkb-T2A-Gal4* resulted in a much broader coverage than expected, possibly due to its early regional gap gene expression (Figure 3.3D). Finally, *vnd-*, *ind-*, and *msh-T2A-Gal4* have immortalized 14, 14-15, and 8-9 NBs each (Figure 3.3E-G). Interestingly, the original D/V gradient of

the three columnar genes is now transformed into a medial/lateral NB distribution.

To visualize the corresponding progeny of these NBs into the adult stage, the initial T2A-Gal4 KI/*dpn* intersection has to be preserved further, intuitively by triggering an additional irreversible event. Two new recombinases and their recombination targets, KD/*KDRT* and CRE/*loxP*, were hence introduced for sequential excisions of stop cassettes for finer transfer of activities (Siegal and Hartl 1996; Nern *et al.* 2011), and FLP/*FRT* was intentionally saved for future stochastic sampling (see below). In this new immortalization scheme, a T2A-Gal4 KI first removes the *KDRT* stop cassette between a *dpn* promoter and a CRE recombinase by expressing UAS-KD. CRE in turn removes the *loxP* stop cassette between an actin promoter and a LexA::p65 driver to create a permanent driver, *act-loxP-LexA::p65*. As a result, by restricting the original T2A-Gal4 activity through the *dpn* promoter to certain NBs, and preserving it further through the actin promoter, a lineage-restricted driver is created to present all of the adult lineages derived from a specific molecular marker of interest (Figure 3.4A).

The adult lineage coverage is consistent with the one made at WD larval stage for each of the six markers (Figure 3.4B-G). *lab* and *unpg* both correspond to two cerebral lineages, while *hkb*+ lineages continue to be a massive collection. The three columnar genes generate about 10 lineages each; however, the original larval medial/lateral NB distribution becomes obscure in the adult brain.

Additionally, it is noteworthy that *msh*+ lineages densely innervate the AL, a known deutocerebral neuropil, which agrees with the major expression of *msh* in deutocerebral NBs (Figure 3.4G) (Urbach and Technau 2003b).

Mapping individual *lab*+, *unpg*+, *vnd*+, *ind*+, and *msh*+ lineages

After acquiring the rudimentary knowledge of the adult lineage coverage derived from each of the six molecular markers, we were enthusiastic to know their individual identities. Therefore, an extra sampling strategy is necessary under the general immortalization scheme. This was achieved by bringing in *hs-FLP* and *LexAop2>FRT-stop-FRT>myr::tdTomato*. Heat shock induced expression of FLP at early embryonic stage could stochastically flip-out the *FRT* stop cassette in random NBs. Afterwards, the lineage-restricted *act-loxP-LexA::p65* driver can turn on *lexAop-FRT-myr::tdTomato* and label lineages that experienced such flip-out events with an additional reporter, tdTomato (Figure 3.5A).

Both *lab* and *unpg* cover their own one unique AL lineage, ALlv1 and ALv1 respectively (Figure 3.5B). Also, they both contain the WEDa1 lineage (Figure 3.5B), suggesting co-expression of these two markers in the corresponding embryonic NB. The lack of appearance of *lab*+ lineages along the SPG and SEG border again indicated the demise of the majority of tritocerebral lineages during development. We did not pursue the sampling of *hkb*+ lineages because *hkb* as a gap gene apparently has a much broader expression before NB delamination, during which Urbach and Technau performed their survey. Immortalization with

hkb-T2A-Gal4 consequently produced a huge collection of adult lineages. Hence, we expected there exist no straightforward correspondence between our immortalization strategy and their *hkb*+ NB mapping in this specific case.

vnd+ lineages are exhaustively sampled. There are 16 of them: SMPad1, SMPp&v1, VLPa2, CREa1, CREa2, AOTUv1, AOTUv3, AOTUv4, ALv1 (*unpg*+), LALv1, VESa1, VESa2, FLAa1, FLAa2, FLAa3, and WEDd1 (Figure 3.6). This coverage is identical to an independent parallel analysis made by our lineage team with the same *vnd-T2A-Gal4* driver (Ying-Jou Lee, unpublished data).

We currently have identified 13 *ind*+ lineages: LHa1, LHl2, SLPal5, SLPav1, SLPav3, SLPpl1, VLPd1, VLPp1, VPNp1, ALv1 (*unpg*+ and *vnd*+), LALv1 (*vnd*+), VESa1 (*vnd*+), WEDa1 (*lab*+ and *unpg*+). The large collection of LH and SLP lineages contributes to the prominent labeling of dorsal neuropils under general immortalization, a feature consistent with the fact that *ind* is largely expressed in protocerebrum. We expect to discover several additional *ind*+ lineages after continual sampling because there are 14-15 NBs immortalized at the WD larval stage (Figure 3.3F).

There are 11 *msh*+ lineages: LHl1, SLPal1, VLPl4, VLPp1, VPNp1 (*ind*+), ALad1, ALl1, ALlv1 (*lab*+), ALv2, VESa1 (*vnd*+ and *ind*+), and PSa1 (Figure 3.8). Amazingly, 4 out of 5 AL lineages are included in the *msh*+ lineages. Since AL is an important olfactory information processing and relaying neuropil, we believe knocking-in different effectors in *msh* will be a desirable means of sophisticated genetic intersection for functional AL studies in the future.

We summarize the lineage coverage of each of the six T2A-Gal4s we created (Table 3.1). Interestingly, the three columnar genes together only cover one third of the cerebral lineages (34/106), which again suggests DV pattering does not play a major role in cerebrum development as in VNC. *vnd*+, *ind*+, and *msh*+ lineages are broadly distributed over the cerebrum (Figure 3.6, 3.7, 3.8). Therefore, the spatial organization becomes more obscure even though the initial DV patterning is roughly preserved in tritocerebrum and deutocerebrum.

Also, with proper intersection design, we have clues to target 7 lineages (VESa1, WEDa1, VLPp1, ALlv1, ALv1, LALv1, and VPNp1) individually. Being the intermediate columnar gene, *ind* corresponds to lineages that overlap with *vnd*+ lineages (VESa1, ALv1, and LALv1) and *msh*+ lineages (VESa1, VLPp1, and VPNp1), while *vnd* and *msh* only produce one common lineage, VESa1, which is also covered by *ind*. The significance of the overlap remains to be addressed since the three columnar genes are mostly mutually exclusive in both pNR and vNR. Other than co-existing in the same NB, sequential expression of the columnar genes can be the reason for the overlapping in immortalization coverage. Our studies thus revealed new aspects of the interplay between early patterning genes for future mechanistic studies.

Discussion

In general, we are pleased with the fruitful lineage analyses made only possible by our valuable T2A-Gal KI drivers and the insights provided by Urbach and Technau's NB map. Our irreversible immortalization design revealed the fact that early patterning genes are rigidly regulated and their expression possibly becomes more and more restricted within their own initial NB coverage. Otherwise, we would have obtained a bigger collection of lineages for each of the molecular markers with recovery of small lineage clones that come from immortalization at later developmental stages. Nevertheless, there exists obvious inconsistency between these two works that we need to mention and discuss its possible sources.

First, the tritocerebral lineages remain mysterious. Our major attempt to reveal them by generating *lab-T2A-Gal4* led us to believe that most of them do not survive to WD larval and adult stages. Also, *lab* and *unpg* possibly have overlapping activities in the NB that produces the WEDa1 lineage, a candidate of tritocerebral lineages; yet, *unpg* does not have expression in tritocerebral NBs in the earlier reports. One possible explanation is our *dpn* promoter does not cover the tritocerebral NB expressions, and immortalization failed accordingly. We will address this issue in follow-up studies by direct examination of *dpn* promoter activities in tritocerebrum and immortalization with a parallel NB-specific promoter (*asense* promoter). On the other hand, it is believed that three

columnar genes roughly subdivide the tritocerebrum and its NBs into three groups. We did recover lineages situated right above SEG from immortalization of *vnd-T2A-Gal4* (VESa1, VESa2, FLAa1, FLAa2, FLAa3, and WEDd1), *ind-T2A-Gal4* (VESa1 and WEDa1) and *msh-T2A-Gal4* (VESa1 and PSa1). They may be candidates for the tritocerebral lineages.

Second, our lineage coverage does not match up well with Urbach and Technau's NB annotations. We are curious to know the identities of the five NBs that correspond to the five AL lineages. ALv1 is included in *unpg*+, *vnd*+, and *ind*+ lineages. However, *unpg* and *vnd* do not have common NB expression in deutocerebrum, with the closest candidates being Dd1/Dd5 or Dv7/Dv8. Because NB Dd1 is additionally covered by *ind*, we suspect Dd1 produces the adult ALv1 lineage. Also, ALlv1 was discovered in both *lab*+ and *msh*+ lineages, but there is no shared NB in deutocerebrum between *lab*+ (Dv2 and Dv4) and *msh*+ NBs (Dd2 to Dd13). NB Dd5 expresses both *vnd* and *msh*. The only common adult lineage of *vnd*+ and *msh*+ is VESa1, which is also included in *ind*+ lineages; unfortunately, NB Dd5 does not express *ind*. There is no overlap between *vnd*+ and *ind*+ NBs to explain common *vnd*+ and *ind*+ coverage of LALv1; also, there is no overlap between *ind*+ and *msh*+ NBs to predict VLPp1 and VPNp1 either. Lastly, LHl1 and SLPal1 from the *msh*+ lineages are located well in the protocerebrum region, but *msh* does not seem to be expressed in protocerebral NBs.

One other significant discrepancy is the mismatch of NB/Lineage numbers. Leaving tritocerebral NBs aside, there are 12 annotated deutocerebral *msh*+ NBs; yet, we only recovered 11 *msh*+ adult lineages in total after immortalization. Similarly, Urbach and Technau found 17 *vnd*+ deuto- and protocerebral NBs at the embryonic stage, while we recognized 16 lineages in the whole brain after exhaustive sampling. Currently, there are 13 *ind*+ lineages in our collection, but there are only 6 *ind*+ deuto- and protocerebral NBs.

There are several ready explanations for such disagreement between these two studies. First, the endogenous embryonic expression of these molecular markers might be too transient or weak to trigger the first stop cassette excision of the cascade of immortalization; hence, lower numbers of our lineages than their NBs. Second, not all of the embryonic NBs produce lineages that survive into the adult stage, notably the tritocerebral NBs. Third, the Technau group's NB map might not be 100% correct at all times because of the difficulties in consistent embryo sample preparation and precise NB annotation without rigid landmarks. We have this reasonable doubt owing to the conclusion that the likely NB candidates for ALv1 are Dd1/Dd5 and Dv7/Dv8, which arise in proximity in neuroectoderm. Finally, either our T2A-Gal4 KIs or their antibody/reporter mapping does not 100% faithfully represent the endogenous expressions of these molecular markers. In Urbach and Technau's map, we noticed that among the three columnar genes, *ind* was the only one with its expression annotated based on *in situ* hybridization. It is understandable that a gene's expression can

differ significantly over transcription, translation, and post-translation, not to mention we access it through binary induction.

Despite the variance, these T2A-Gal4 KIs are nonetheless invaluable for lineage analyses. With a proper intersection design, we will have the ability to target certain lineages of interests, examples: VESa1, WEDa1, VLPp1, ALlv1, ALv1, LALv1, and VPNp1. As mentioned earlier, the AL lineages are one of the obvious entry materials for detailed mechanistic studies. We are eager to discover the major determinants for their shared features as well as distinct characteristics. Thanks to this study, we realized the three columnar genes are great candidates to initiate this investigation. In a parallel effort led by Zhiyong Liu and Rosa Miyares in the lab, we can specifically target and differentially label four of the five AL lineages (ALad1, ALl1, ALlv1, and ALv2). I will first use publicly available fly reagents to knock-down (UAS-RNAi) or ectopically express columnar genes (UAS-ORF) in AL lineages and evaluate the effects on their elaborate and sensitive morphologies.

One striking observation over comprehensive lineage analyses is the potential discovery of lineage duplication. CREa1 and CREa2 are two *vnd*+ lineages of such a character. Ying-Jou Lee revealed their similar developmental profiles while performing in-depth birth dating analyses (unpublished data). In short, half of the CREa1 lineage is composed of neurons of similar morphological characteristics and temporal birth order to one half of the CREa2 neurons, while both CREa1 and CREa2 produce distinct other halves that separate them from

each other. We are interested in revealing the determining factors among the combinatory codes that lead to the diversification of CREa1 and CREa2. Interestingly, we suspect there is another example pair of lineages carrying such properties: AOTUv3 and AOTUv4. A constantly growing collection of KIs of early patterning genes will allow us to obtain intersections that lead to lineage-specific drivers, which are essential for cell fate studies.

Finally, we want to address one of the shortcomings of these T2A-Gal4 KIs. Unfortunately, some of them accumulate mutations that lead to loss of Gal4 expression over time. We have no way to prevent such naturally occurring undesirable events right now. However, by our continual effort to improve Golic+ with RMCE ability, we hope to gain the flexibility to generate KIs with different modules whenever necessary; hence, it is possible to circumvent this issue in the future.

Materials and Methods

Generating T2A-Gal4 KIs

Homology arms of about 3 kb each were cloned into pTL1 for knocking-in T2A-Gal4 in *lab*, *unpg*, *hkb*, *vnd*, and *ind*. T2A-Gal4 CDS was inserted right after the second last amino acid codon for each of these five genes. Details of *msh-T2A-Gal4* KI cloning can be found in Material and Methods, Chapter II. The following primers were used to amplify the homology arms:

lab_55AgeI: TACGACCGGTGGACAAGACTTGGGGTTTTACG;
lab_53StuI: AAGGCCTACTTTGCTTGCTCGTGGG;
lab_35EcoNI: TACGCCTTTCGAAGGTGAAGGGCTTTCTTCGATGTTG;
lab_33MluI: TACGACGCGTGACCAAAAAGGACATTGTGGG;
unpg_55AgeI: TACGACCGGTGATGGCTAACATTCGTTGTTGC;
unpg_53StuI: AAGGCCTATAGATGCTCCTGGCGAGG;
unpg_35EcoNI: TACGCCTTTCGAAGGTGATGGAGTTGGGCTCCT;
unpg_33MluI: TACGACGCGTCAAATTGTTCAAATGACTCCCGC;
hkb_55NgoMIV: TACGGCCGGCCTGCGGAATGCACCTTC;
hkb_53StuI: AAGGCCTGTAGCCATAAAGGTATGAGTACATGG;
hkb_35BamHI: CGGGATCCTGAGAGGGAAATCCTGGCG;
hkb_33EcoNI: TACGCCTTTCGAAGGCGCATTATGGCCAGGAAAGC;
vnd_55AgeI: TACGACCGGTGATCAAGGAGAACGAGCTATACG;
vnd_53StuI: AAGGCCTGGGCCACCAGGCGG;
vnd_35PmeI: TACGGTTTAAACTAATATTGCTAGGAACTGGCATTCAC;
vnd_33MluI: AAGTACGCGTAACTGGAATAAGTTC;
ind_55AgeI: TACGACCGGTGAAGTCTAAGTCAATAGCACGATC;
ind_53StuI: AAGGCCTCGCCTCAACCTTCAATTCGTG
ind_35PmeI: TACGGTTTAAACTAGAATCAATCACCTATTAAACCATAAGA;
ind_33MluI: TACGACGCGTGCGGGACACTCGAAAGTTG.

Donors were integrated in either *attP40* (*lab*, *hkb*, *vnd*, *ind*) or *VK00027* (*unpg*), depending on the locations of the endogenous genes. We used traditional Golic heat shock strategy for donor release; however, lethality

selection was implemented for easy candidate recovery. For details, refer to “Traditional Golic Heat Shock”, Material and Methods, Chapter II.

The following BAC clones were used: BACR69M19 for *lab*; BACR82N17 for *unpg*; BACR30G03 for *hkb*; BACR16G06 for *vnd*; BACR42C08 for *ind*. These primers were used in genomic PCR for T2A-Gal4 KI confirmation:

Gal4-1: CACACGCTTGTTCAATACTACTCAG;
Gal4-2: GATACTCCACCGAACCCAAAGAAG;
unpg_g1: TTGGTTAGACGTTCGACAATCCTTAG;
unpg_g2: CACTTTGTTCGCTGAACGCTCC;
lab_g1: GTGGAAGCGATCCCCGATAATG;
lab_g2: TTCCTTTTGGTTTGGTTTGGTCCC;
hkb_g1: TCGAACAATGCAGTTGGACGTAATC;
hkb_g2: GCCAGGAAAGCACGTTACCATT;
vnd_g1: CTTTGGCGACGAGATGTCCTC;
vnd_g2: GCTGCAGATATGCGAGCATAAGC;
ind_g1: GAAGCCGATGCTATCTTGGAGATAAG;
ind_g2: CTTTAAACTGAAGAGGTTCTGCACTTC.

Fly Strains

Transgenes constructs were inserted into various attP sites in *Drosophila* genome. The T2A-Gal4s, generated by gene targeting, used in this study include: *lab-T2A-Gal4*, *unpg-T2A-Gal4*, *hkb-T2A-Gal4*, *vnd-T2A-Gal4*, *ind-T2A-Gal4*, and *msh-T2A-Gal4*. Transgene used for characterizing the dynamics of various T2A-Gal4 activities at different developmental stages was *UAS-GFP::Dbox* in *VK00027*. Transgenes used for immortalization of the T2A-Gal4 activities through a *dpn* NB enhancer during larvae development include: *UAS-FLP* in *attP18*, *dpn>FRT-stop-FRT>LexA::p65* in *su(Hw)attP8*, *LexAop2-myr::GFP* in *attP40* plus *su(Hw)attP5*. Transgenes used for immortalizing T2A-

Gal4 pattern through a *dpn* NB enhancer to lineage-restricted activity include: *dpn<KDRT-stop-KDRT<Cre::PEST* in *su(Hw)attP8*, *act^loxP-stop-loxP^LexA::p65* in *attP40*, *LexAop2-myr:GFP* in *su(Hw)attP5*, *UAS-KD* in *VK00027*. Transgenes used for sampling the immortalized T2A-Gal4 pattern through a *dpn* NB enhancer to lineage-restricted activity include: *hs-opFLP* in *attP18*, *dpn<KDRT-stop-KDRT<Cre::PEST* in *su(Hw)attP8*, *act^loxP-stop-loxP^LexA::p65* in *attP40*, *LexAop2-myr:GFP* in *su(Hw)attP5*, *UAS-KD* in *VK00027*, *LexAop2>FRT-stop-FRT>mry::tdTomato* in *su(Hw)attP5* plus *attP2*.

Lineage immortalization and sampling

Larvae and adults with the genotype for lineage immortalization were raised in standard conditions and dissected at corresponding stages. For sampling, eggs were collected every three hours, aged for another three hours, and underwent heat shock in 37°C water bath for different periods of time (20 min for *lab* and *unpg*; 15 min for *msh*; 10-12 min for *vnd* and *ind*) owing to the number of immortalized lineages they contained.

Immunohistochemistry and confocal microscopy

Procedures for embryo preparation, fixation, and staining can be found in Materials and Methods, Chapter II. For embryonic vNR studies, overnight-collected embryos were fixed and stained as previously described (Lacin *et al.* 2009). Larval brains at corresponding stages and adult brains from 3- to 4-day-

old flies were dissected in 1X PBS and fixed for 20 min in 4% paraformaldehyde fixation solution. After fixation, brains were first rinsed 3 times, then washed 3 times (10, 20, 30 min) in PBST (0.5% Triton X-100 in 1X PBS), and finally incubated with primary antibodies (diluted in PBST + 5% normal goat serum) overnight at 4°C. Next day, samples were incubated at room temperature for an additional hour before rinsed 3 times and washed 3 times (10, 20, 30 min) in PBST. Brains were then incubated with secondary antibodies (diluted in PBST + 5% normal goat serum) overnight at 4°C. Finally, they were placed in room temperature for an hour before rinsed 3 times and washed 3 times in an hour. For mounting, brains were transferred on glass slides rinsed 2 times with 1X PBS, rinsed again in SlowFade Gold antifade reagent (InVitrogen) and soaked in SlowFade for fluorescence imaging. Embryo samples were collected, fixed, and stained in the same procedures as the ones in Chapter II.

The following primary antibodies were used in this study: guinea pig anti-Deadpan, 1:1000 (kindly provided by James B. Skeath); rabbit anti-Msh, 1:500 (kindly provided by Chris Q. Doe; ISSHIKI *et al.* 1997); rat monoclonal anti-GFP, 1:1000 (NACALAI TESQUE, INC. 04404-84); rabbit polyclonal anti-GFP, 1:1500 (Invitrogen #A-11122); rabbit polyclonal anti-DsRed, 1:500 (Clontech, Living Colors, 632496); mouse anti-Bruchpilot, nc82 mAb, 1:100 (DSHB). Secondary antibodies used include Alexa Fluor 488 goat anti-rat IgG, 1:500 (Invitrogen #A-11006), Cy3-conjugated goat anti-rabbit antibody, 1:200 (Jackson ImmunoResearch #111-165-144), and Cy5-conjugated goat anti-mouse

antibody, 1:200 (Jackson ImmunoResearch #115-605-146). For embryos, Cy3-conjugated goat anti-rabbit (1:200; Jackson ImmunoResearch, 111-165-144) and Hoechst 33342 (1:1000; invitrogen) were used.

Fluorescent signals of whole mount embryos, larval and adult brains were collected by confocal serial scanning using Carl Zeiss LSM710 microscope. Images were processed with Fiji then rotated and cropped with Keynote.

Table 3.1. Summary of lineage immortalization.

Lineages	*lab*	*unpg*	*vnd*	*ind*	*msh*
Unique Lineages for a certain molecular marker			SMPad1 SMPp&v1 VLPa2 CREa1 CREa2 AOTUv1 AOTUv3 AOTUv4 VESa2 FLAa1 FLAa2 FLAa3 WEDd1	LHa1 LHl2 SLPal5 SLPav1 SLPav3 SLPpl1 VLPd1	LHl1 SLPal1 VLPl4 ALad1 ALl1 Alv2 PSa1
VESa1			v	v	v
WEDa1	v	v		v	
VLPp1				v	v
ALlv1	v				v
ALv1		v	v	v	
LALv1			v	v	
VPNp1				v	v

Figure 3.1. Expressions of T2A-Gal4 KIs at embryonic and larval stages.

We used UAS-GFP::Dbox to evaluate our T2A-Gal4 GT designs. All 6 T2A-Gal4 KIs faithfully capture and represent the endogenous expressions of their corresponding targeted genes as compared to the publicly available in situ data from BDGP (http://insitu.fruitfly.org). Interestingly, we also noticed that these genes continue to be selectively expressed in certain NBs (cells with significant larger sizes) through larval stages. Magenta: GFP::Dbox; blue: DNA, Hoechst 33342 staining.

Figure 3.1. Expressions of T2A-Gal4 KIs at embryonic and larval stages.

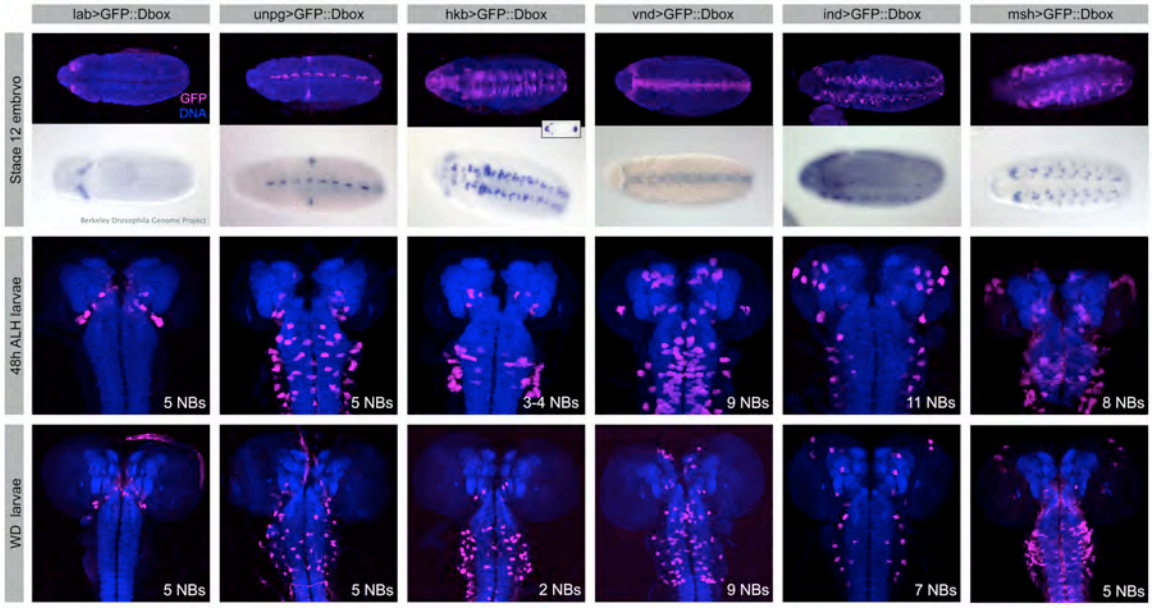


Figure 3.2. *vnd-T2A-Gal4*, *ind-T2A-Gal4*, and *msh-T2A-Gal4* drive reporter expression in medial, intermediate, and lateral NB columns, respectively.

(A-C) Reporter expression (green) and Dpn labeled NBs in abdominal segments of stage 11 embryonic nerve cord are shown. (A) *vnd-T2A-Gal4* drives the reporter expression in the medially located ventral NBs; (B) *ind-T2A-Gal4* drives the reporter in the intermediate columns of NBs; (C) *msh-T2A-Gal4* drives the reporter expression in the lateral column of NBs. (D-E) Three columns of NBs are visible on each side of the midline of early stage 9 embryonic nerve cords that are stained against Msh protein (red). (D) GFP expression driven by *vnd-T2A-Gal4* is detected in medial column NBs (1-1, MP2, 5-2, and 7-1; arrows). (E) GFP expression driven by *ind-T2A-Gal4* is detected in intermediate column NBs (3-2, 4-2, and5-3; arrows). (F) At late stage 11 embryos, *msh-T2A-Gal4* labels NBs 2-4, 4-3, 5-4, and 7-4 (arrows) in addition to several others (not visible at this local plane) White bar marks the midline. Dashed lines mark the segment boundaries.

Figure 3.2. *vnd-T2A-Gal4*, *ind-T2A-Gal4*, and *msh-T2A-Gal4* drive reporter expression in medial, intermediate, and lateral NB columns, respectively.

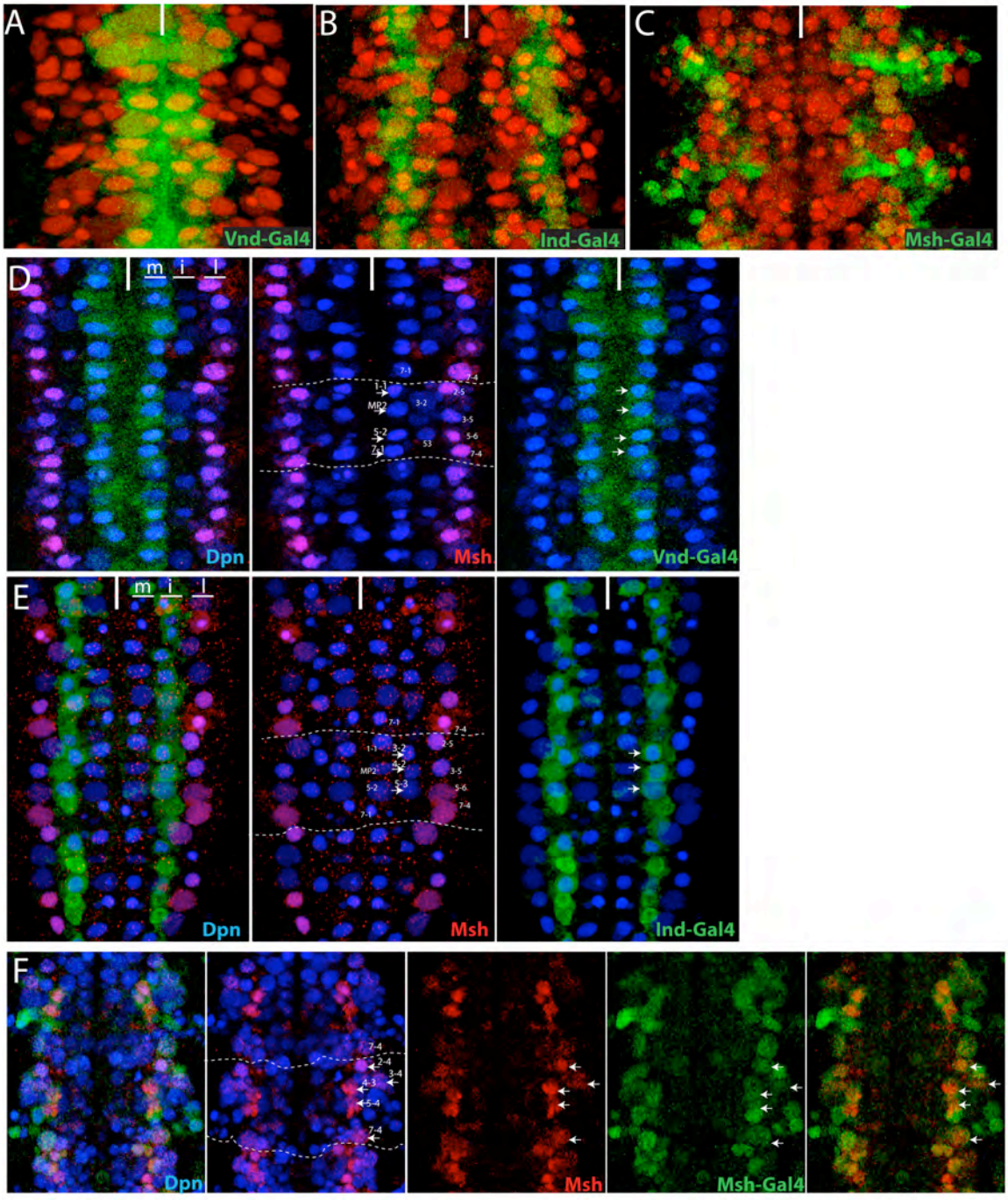


Figure 3.3. Immortalization at the WD larval stage.

(A) An illustration to depict the genetic strategy to immortalize transient expression of molecular markers in NBs into irreversible NB labeling at larval stages. Expressions of a T2A-Gal4 KI in certain NBs result in the removal of a stop cassette between a *dpn* promoter and a LexA::p65 driver. The reconstituted *dpn-FRT-LexA::p65* driver can then turn on myr::GFP to label these NBs continually, plus their immediate progeny cells. (B-G) NBs that express certain molecular markers were revealed accordingly by immortalization at the WD larval stage. Magenta: myr::GFP; blue: neuropil, nc82 staining.

Figure 3.3. Immortalization at the WD larval stage.

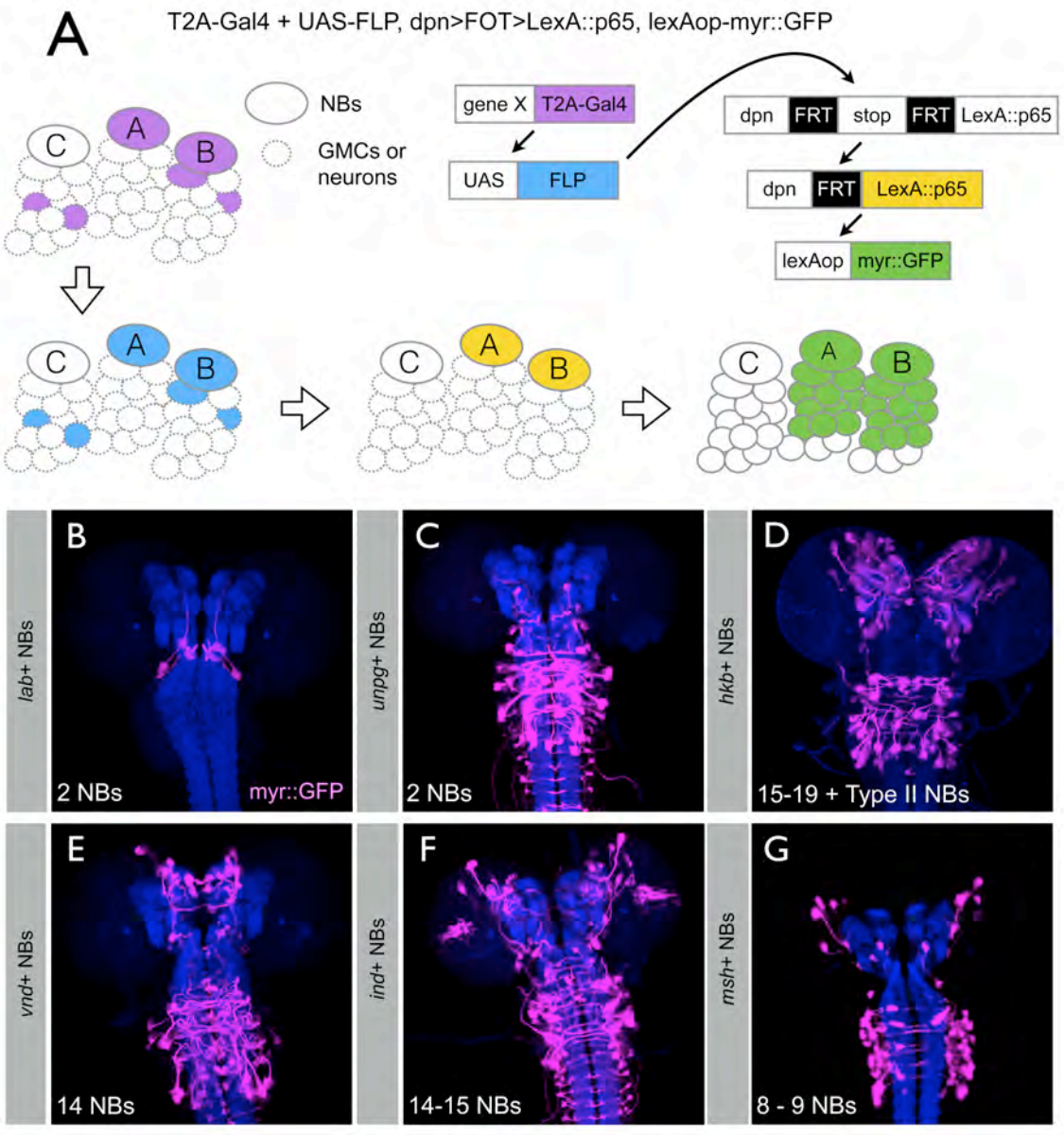


Figure 3.4. Adult lineages of NBs positive for certain molecular markers.

(A) The genetic strategy to reveal lineages produced by NBs that express a certain molecular marker. Two recombinases (KD and Cre) work sequentially to remove two stop cassettes and consequently immortalize early NB expression of a specific molecular marker into irreversible labeling of subsequent progeny cells of the corresponding NBs. (B-G) Lineages of NBs that express certain molecular markers were uncovered by immortalization at the adult stage. Magenta: myr::GFP; blue: neuropil, nc82 staining.

Figure 3.4. Adult lineages of NBs positive for certain molecular markers.

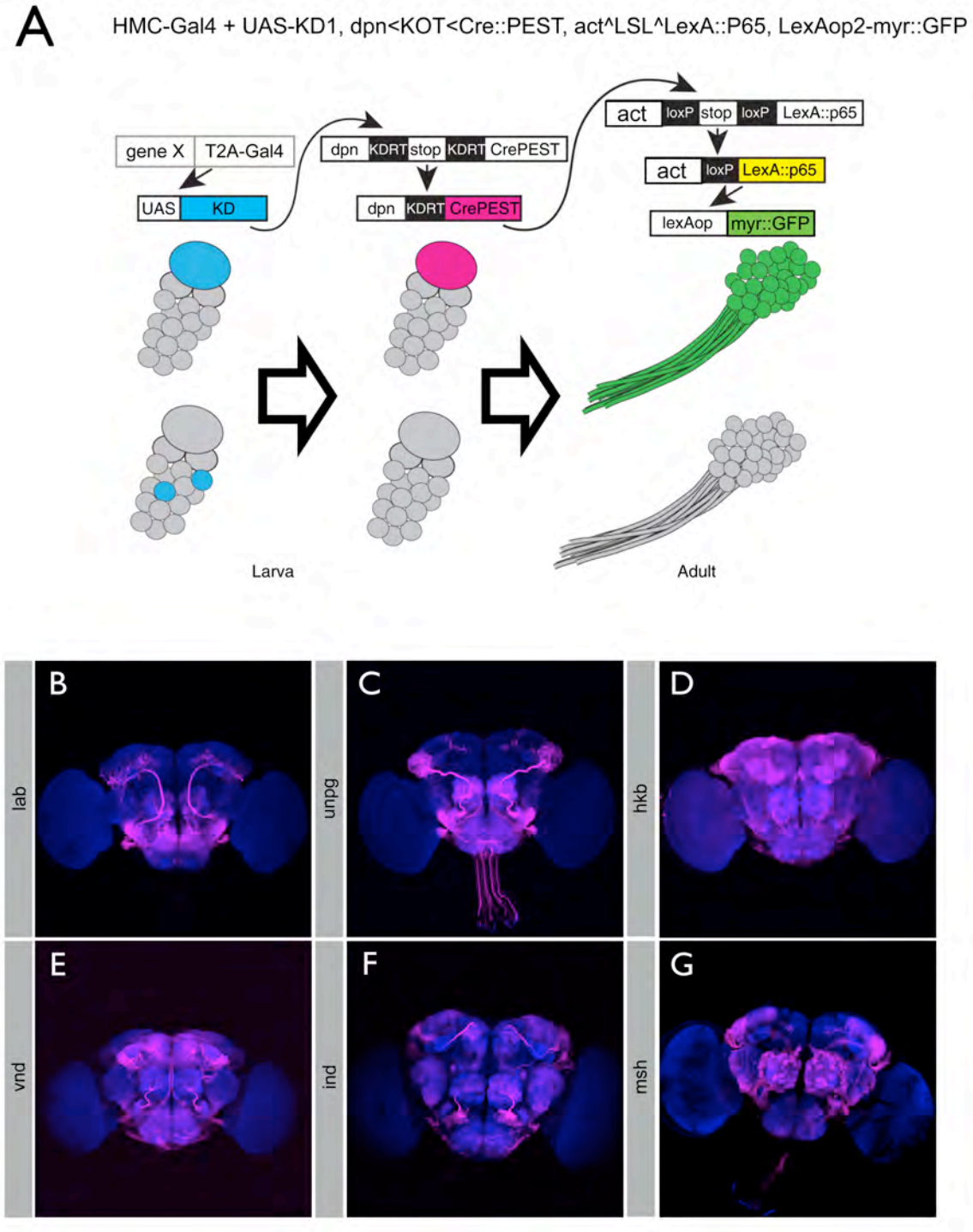


Figure 3.5. Uncovering individual lineages after immortalization.

(A) Lineage sampling was achieved by stochastic removal of an *FRT*-flanked stop cassette in *LexAop2>FRT-stop-FRT>myr::tdTomato* by hs-FLP. Once such random events occur in the immortalized NBs, they will be additionally labeled by myr::tdTomato. (B) *lab*+ lineages (ALlv1 and WEDa1) and *unpg*+ lineages (ALv1 and WEDa1) are distinctly presented with lineage sampling. Magenta: myr::GFP; blue: neuropil, nc82 staining; red: general immortalization, myr::GFP; green: lineage sampling, myr::tdTomato; grey: neuropil, nc82 staining.

Figure 3.5. Uncovering individual lineages after immortalization.

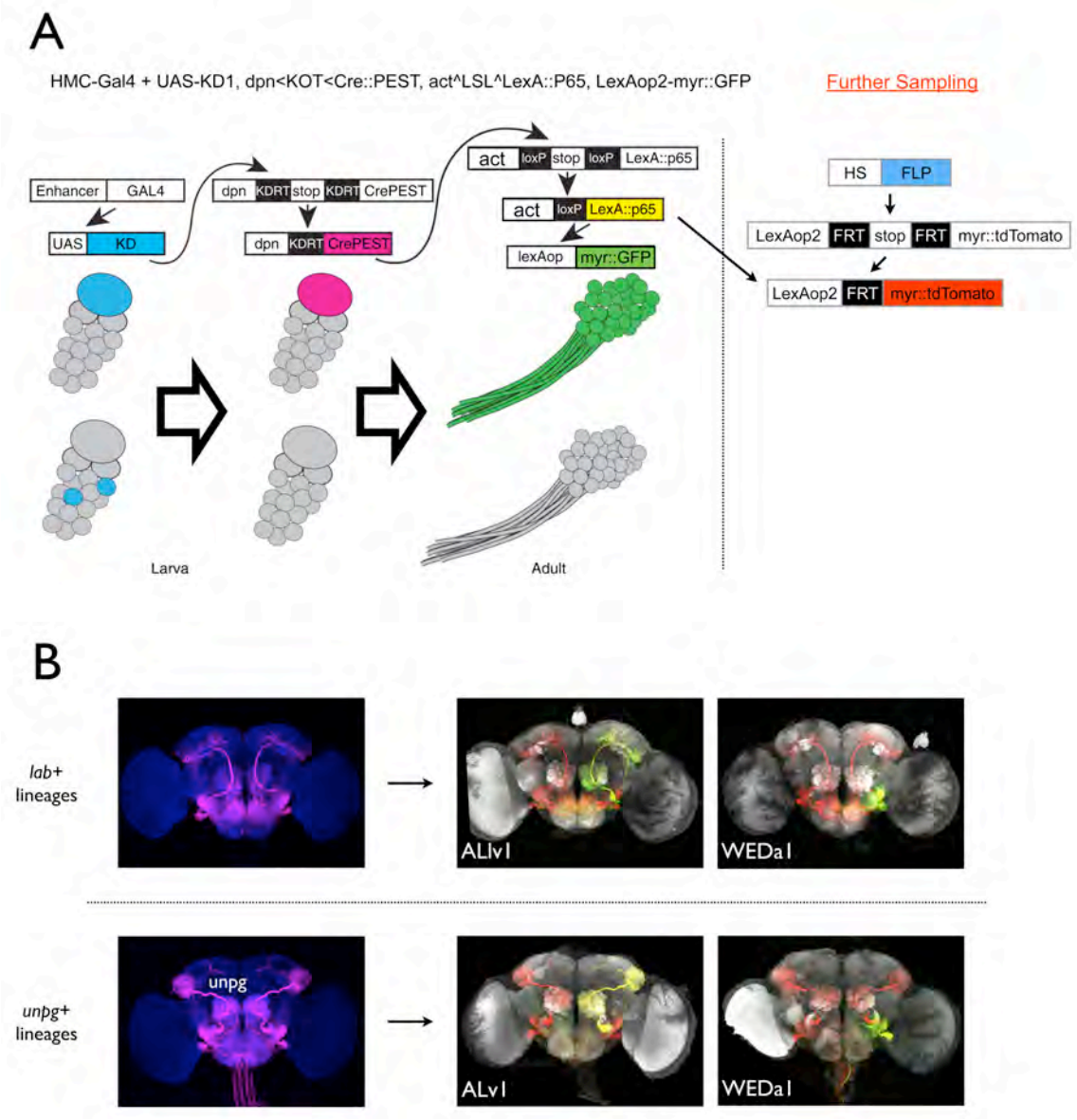


Figure 3.6. *vnd*+ lineages.

(A-P) Representative images of 16 *vnd*+ lineages are presented here after exhaustive sampling. In cases of multiple lineage hits in a single brain, the target lineages are circled with white dashed lines. Red: general immortalization, myr::GFP; green: lineage sampling, myr::tdTomato; grey: neuropil, nc82 staining.

Figure 3.6. *vnd*+ lineages.

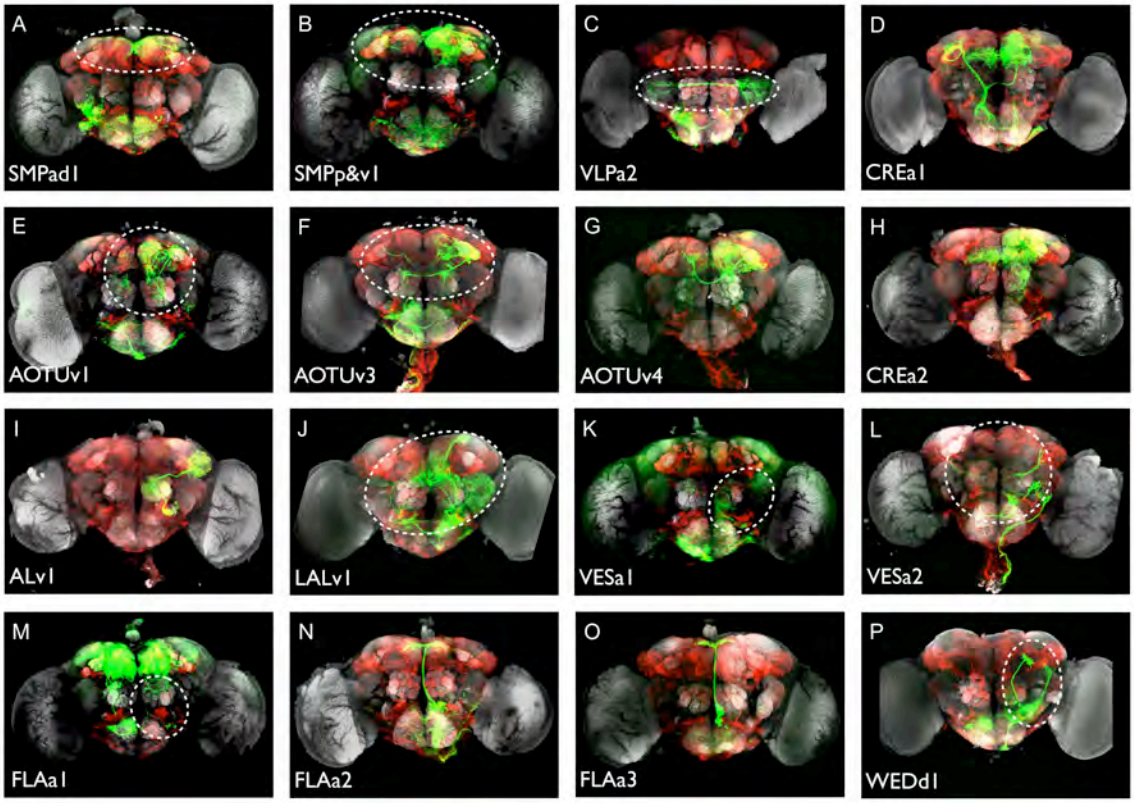


Figure 3.7. *ind*+ lineages.

(A-M) Representative images of 13 *ind*+ lineages are presented here. In cases of multiple lineage hits in a single brain, the target lineages are circled with white dashed lines. Red: general immortalization, myr::GFP; green: lineage sampling, myr::tdTomato; grey: neuropil, nc82 staining.

Figure 3.7. *ind*+ lineages.

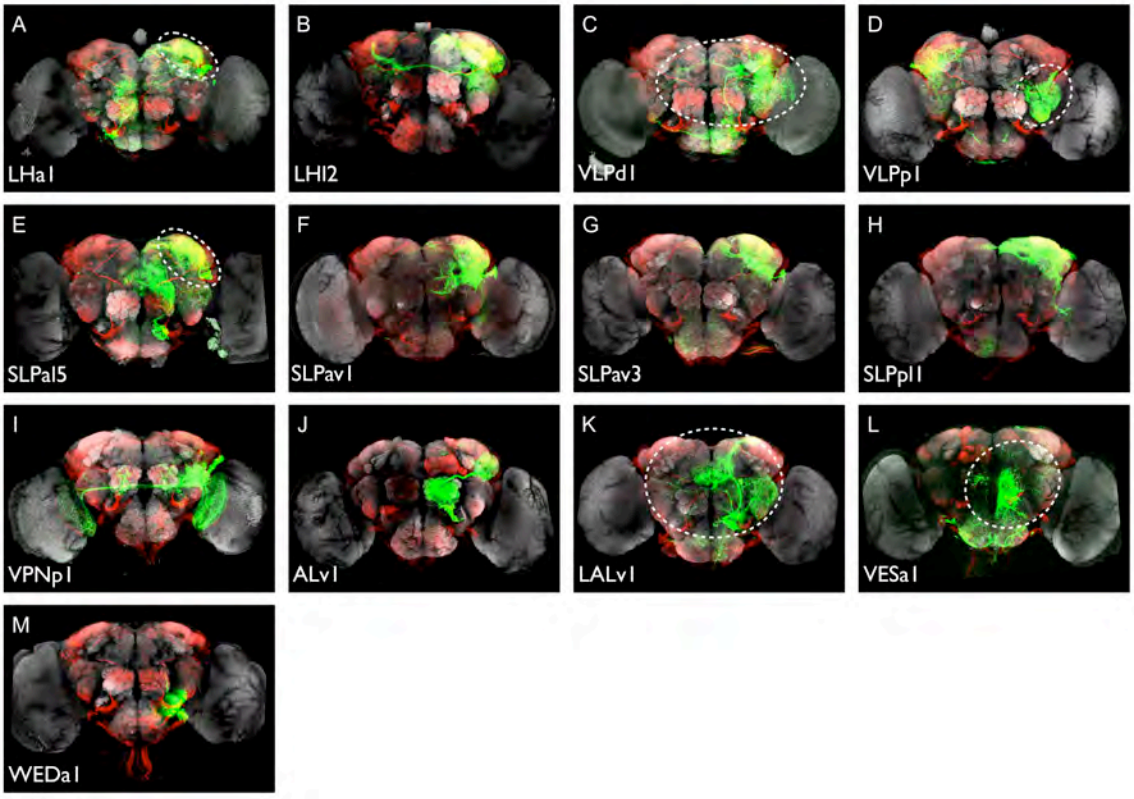
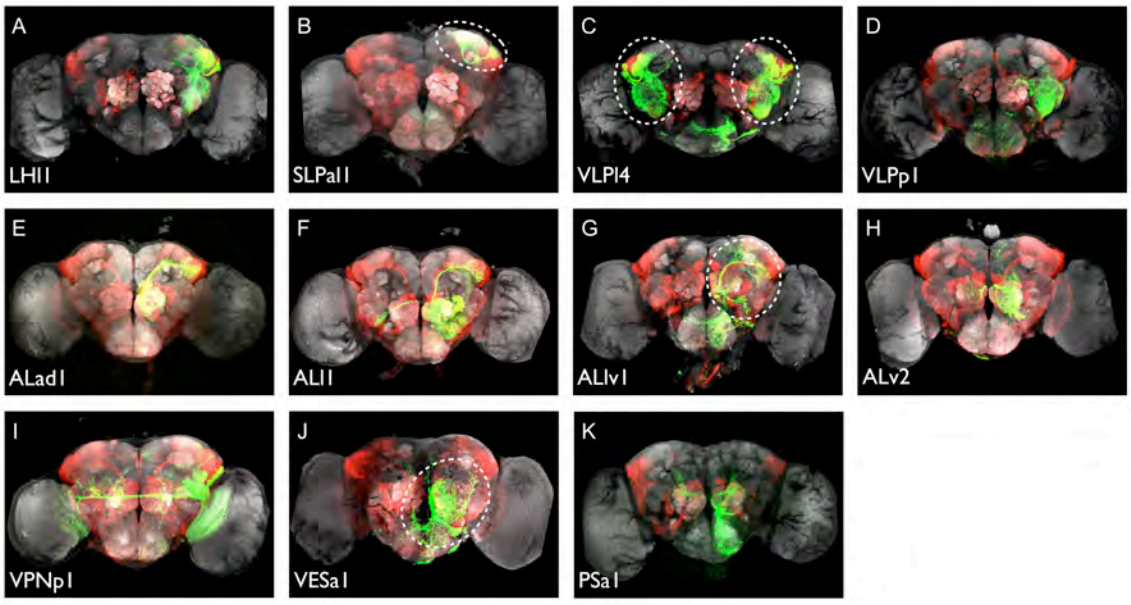


Figure 3.8. *msh*+ lineages.

(A) Representative images of 11 *msh*+ lineages are presented here after exhaustive sampling. In cases of multiple lineage hits in a single brain, the target lineages are circled with white dashed lines. Red: general immortalization, myr::GFP; green: lineage sampling, myr::tdTomato; grey: neuropil, nc82 staining.

Figure 3.8. *msh*+ lineages.



CHAPTER IV

Conclusion

Dissecting the mechanisms of brain development remains one of the most challenging and attractive topics in developmental biology. It never stops amazing me how a number of cells specified from a two dimensional layer produce a huge collection of diversified progeny that in turn elaborate and intertwine to form a structurally complex and functionally sophisticated brain.

Drosophila melanogaster serves as one of the most useful and popular model organisms for development studies due to numerous historical and economical reasons. Mostly, its biology is relatively simple yet complex enough for us to obtain valuable insights for mechanisms evolutionally conserved across species. Also, what comes with it is an almost unmatchable set of genetic tools that help us introduce precise and effective manipulations that are essential for deducing cause-and-effect relationships. Furthermore, its famous stereotypic brain lineages and neuronal elaborations further facilitate the characterization and analyses of individual lineages and corresponding neurons in different experimental settings.

But, still, why do we bother to continually invest significant energy in studying this seemingly distant relative? If the famous discovery of embryonic body plan by Edward Lewis, Christiane Nüsslein-Volhard and Eric Wieschaus is

not convincing enough, how about the amazing similarity in our olfactory systems, thanks to the beautiful work pioneered by Richard Axel and Linda B. Buck? Not to mention a variety of behaviors a little fly can demonstrate: learning and memory, courtship, sleep circadian rhythms, aggression, addiction, and other social behaviors. Studying them did and will continue to provide insights that couldn't be easily acquired in "higher" organisms.

As for the fly brain development, just recently a mammalian type proliferation pattern was discovered in fly, which produces the type II (or PAN) lineages. By depositing a series of transient amplifying precursors that proliferate like short-lived NBs, this new proliferation strategy significantly increases the number of progeny cells a NB can produce. More importantly, all 8 of the type II lineages generate neurons that together innervate more broadly in an adult brain, much different from the relatively regional type I lineages.

Additionally, there are 4 MB lineages standing out among the type I lineages with the longest proliferation windows and without dormancy during embryo/larva transition. They produce fairly large number of neurons yet with fewer morphologically distinguishable types. It was further found that MB lineages have intrinsic and plastic developmental programs. They can uncouple their NB cycling from temporal fate transitions that are closely linked to extrinsic cues directed by overall organismal growth and development. Therefore, external perturbations, such as starvation, can lead to drastic changes in the numbers of MB neurons.

It remains to be seen what are the determining factors for the creation of such novelties among the brain lineages. What is transforming type II lineages to deposit more progeny with more diversity and broader coverage in roughly the same amount of time? Is there a reason for MB, an information processor, to be released from the common developmental constraints? Understanding the sources of their divergence should provide us new perspectives on how complexity and plasticity contribute to possibly higher brain functions in parallel to the general rigid development and likely basal activities. Additionally, we hope to find the patterning cues that uniquely belong to either one of the NBs of the CREa1/CREa2 or the AOTUv3/AOTUv4 lineages. Each NB pair might have their NBs situated so close that only few patterning cues are differential. Identifying such cues will be significant breakthrough of both developmental and evolutionary biology.

We plan to unveil the mystery of fly brain development by a bottom-up approach, step by step, tracing the process from its origin and making sense of this beautiful interplay of molecules, cells, and network. Generally, owing to fruitful lessons learned from VNC studies, we believe *Drosophila* CNS is laid out by three fundamental mechanisms: (i) NBs are first specified by early positional cues and acquire their own identities, (ii) each NB then deposits a series of unique progeny cells that are patterned mostly by intrinsic temporal codes, (iii) environmental factors and organismal experiences can further bring in plasticity in brain development.

However, much less is understood about the development of *Drosophila* cerebrum because of its higher complexity and obscure organization. Although vNR has a more comprehensible grid-like patterning logic, it only gives flies a simpler and repetitive VNC structure. Regarding the fly brain, divergence from the original developmental plans again leads to complexity that achieves a more sophisticated organization during evolution. Obscure patterning is not illogical. Though challenging, it simply is a logic we yet understand and desperately need to study to comprehend how a real complex structure for advanced functions is built from delicate variations in the beginning.

Mechanistic studies in fly brain are often hindered by the lack of systematical ways to introduce effective manipulations in cells with common pedigrees. In fact, other than the distinct MB lineages, thorough and unambiguous tracking of lineages from embryonic to adult stages has never been achieved before. Without continual tracing, it is unreasonable to make convincing and substantial interpretations of any perturbations introduced.

Recognizing the necessity to break through such a bottleneck and make the embryo/adult connections, we embarked on developing novel strategies to generate new “drivers” that can help us follow and manipulate lineages. Our work was inspired by Urbach and Technau’s effort to annotate expression of numerous “molecular markers” in all of the 106 cerebral NBs. Their molecular marker/NB map is the first clue to finding the determinants of NBs’ individual identities. What we need next is a way to perform reverse genetics based on this

information and systematically address molecular markers' roles in the NB specification.

Luckily, Awasaki *et al*. pioneered the method to immortalize cerebral lineages from the embryonic stage with the help of a pan-NB *dpn* promoter. Accordingly, we decided to create T2A-Gal4 KIs for these molecular markers so that we can link molecular markers, NBs, and lineages together using the immortalization technology. Unfortunately, genome editing in fly wasn't efficient enough for large-scale studies when we initiated our project. Hence, we patiently started with improving the traditional Golic gene targeting.

Abandoning the traditional Golic heat shock induction at larval stage, our proposal was to hijack adult ovaries to continually provide donor DNA in each CB so that the laborious effort of heat shock incubation or embryo microinjection can be bypassed with simple fly breeding. We identified and characterized the CB-specific *bamP* promoters for this purpose. Additionally, we implemented a lethality selection to further help us save the energy for candidate screening. However, we found that HR does not occur naturally in CB, or happens in a much lower rate than HR in the larval primordial germ cells. Only after we further introduced sequence-specific DSBs by the CRISPR/Cas technology, the adult ovaries became a feasible tissue for GT. We named our GT package Golic+ (**g**ene targeting during **o**ogenesis with **l**ethality **i**nhibitor and **C**RISPR/Cas plus) to indicate its various features and to honor the pioneering work of the Golic group (Rong and Golic 2000; Gong and Golic 2003).

In general, Golic+ performs at a comparable efficiency for HR as the currently established direct embryonic injection strategy. However, Golic+ does require more time for its requirement of a starting {donor, gRNA} transformant if one round of task of either strategy is enough for recovering sufficient numbers of GT candidates. Also, direct injection provides flexibility that the transgene-based Golic+ can not match, such as multiple gRNAs, double nicking, large deletion, single-stranded oligonucleotide for small modification, etc. On the other hand, when it comes to low-efficiency target loci, we believe Golic+ surpasses direct injection by effortlessly providing unlimited trials from basic breeding because it extensively employs fly genetic designs to save us the labor of microinjection and screening. We believe it remains to be seen which strategy will win out, but Golic+ is a well-thought alternative that has great potentials. Researchers shall choose a strategy based on the scale of their projects and the nature of GT alleles they intend to create.

In any case, we will continue to improve Golic+ in multiple directions to make it more appealing for the fly community. First, we are interested in investigating the high GT efficiency and low rate or complete absence of false-positives resulting from using circular donor DNA. We have generated *bamP-CF* that will ideally release uncut donor plus CRISPR/Cas-induced DSB in every CB. Second, we wish to exclude escapers from further chromosomal mapping by pre-selecting red-fluorescent-eyed candidates (*3X-riTS-Rac1*^{V12} with *3xP3-RFP*) that must carry the repressor, *lexAop-rCD2i*. Third, by incorporating an additional

integration system (BxbI/attPX) we intend to adopt the RMCE concept for versatile and economical retargeting. Finally, we are eager to explore the possibility of assembling a screening platform with a constant supply of targeted mutants by exploiting CRISPR/Cas in cystoblast. Using *bamP-Cas9* and other necessary transgenes, we wish to fully exploit adult ovaries as mutagenesis pipelines.

We generated 6 molecular marker T2A-Gal4 KIs accordingly with our own GT strategies. They faithfully captured the endogenous expression of the target genes and resulted in insightful lineage immortalizations. Also, each T2A-Gal4 KI brought us either comparable or slightly smaller number of immortalized lineages compared to Urbach and Techanu's initial NB annotations. We believe that these patterning genes are restrictively expressed within their initial coverage, possibly in fewer and fewer NBs judging from our direct T2A-Gal4 expression data (Figure 3.1). This conclusion is also supported by our constant recovery of full lineage clones, which means immortalization at the embryonic stage, and no clone induction was triggered by these KIs afterwards outside of the original coverage.

There exists significant inconsistency between the results of our study and Urbach and Technau's annotation that, we believe, comes from the limitations of both strategies. Neither expression at the mRNA nor the protein level, determined by *in situ* hybridization or antibody staining, can fully represent a gene's expression, let alone reporting with a secondary reporter through binary induction. T2A-Gal4 definitely bears the inherent flaws of a temporal delay and

amplification in magnitude while representing a gene's expression. However, it is one of the most ideal tools we can generate to transform information into real handles on a cell population of our interest. One resolution to the incompatibility is to annotate our T2A-Gal4s in the same embryo flat preparation, possibly through collaboration considering the skill, experience, and knowledge involved. Yet, we believe the current coarse correspondence is a sufficient starting point because immortalization initiated by the T2A-Gal4s most likely captures the prominent core expression profiles. All in all, the continual intersection of these two work will surly lead to better understandings of the NB fate determination, neuron identity acquisition, and full lineage development.

We hope, equipped with our own efficient and soon-to-be versatile Golic+, the sophisticated immortalization technology, and Urbach and Technau's NB map, we just initiated a new chapter of mechanistic studies in *Drosophila* cerebrum development. Drivers we create down the road can also benefit other functional and behavioral studies. We are determined to reach finer and finer dissection of the neuronal network, and hopefully, one day, come to a better appreciation of this structurally complex and temporally prolonged process of constructing a delicate functioning brain.

References

AKAM, M., 1987 The molecular basis for metameric pattern in the *Drosophila* embryo. Development **101:** 1-22.

ARTAVANIS-TSAKONAS, S., and M. A. MUSKAVITCH, 2010 Chapter one-notch: the past, the present, and the future. Current topics in developmental biology **92:** 1-29.

ARTAVANIS-TSAKONAS, S., and P. SIMPSON, 1991 Choosing a cell fate: a view from the Notch locus. Trends in Genetics **7:** 403-408.

AWASAKI, T., C. F. KAO, Y. J. LEE, C. P. YANG, Y. HUANG *et al.*, 2014 Making *Drosophila* lineage-restricted drivers via patterned recombination in neuroblasts. Nat Neurosci **17:** 631-637.

BAENA-LOPEZ, L. A., C. ALEXANDRE, A. MITCHELL, L. PASAKARNIS and J. P. VINCENT, 2013 Accelerated homologous recombination and subsequent genome modification in *Drosophila*. Development **140:** 4818-4825.

BALLY-CUIF, L., and M. WASSEF, 1995 Determination events in the nervous system of the vertebrate embryo. Current opinion in genetics & development **5:** 450-458.

BASSETT, A. R., and J. L. LIU, 2014 CRISPR/Cas9 and genome editing in *Drosophila*. J Genet Genomics **41:** 7-19.

BASSETT, A. R., C. TIBBIT, C. P. PONTING and J. L. LIU, 2013 Highly efficient targeted mutagenesis of *Drosophila* with the CRISPR/Cas9 system. Cell Rep **4:** 220-228.

BAYRAKTAR, O. A., and C. Q. DOE, 2013 Combinatorial temporal patterning in progenitors expands neural diversity. Nature **498:** 449-455.

BELLO, B. C., N. IZERGINA, E. CAUSSINUS and H. REICHERT, 2008 Amplification of neural stem cell proliferation by intermediate progenitor cells in *Drosophila* brain development. Neural Dev **3:** 8104-8103.

BERTRAND, N., D. S. CASTRO and F. GUILLEMOT, 2002 Proneural genes and the specification of neural cell types. Nature Reviews Neuroscience **3:** 517-530.

BEUMER, K. J., J. K. TRAUTMAN, A. BOZAS, J.-L. LIU, J. RUTTER *et al.*, 2008 Efficient gene targeting in *Drosophila* by direct embryo injection with zinc-finger nucleases. Proceedings of the National Academy of Sciences **105:** 19821-19826.

BEUMER, K. J., J. K. TRAUTMAN, K. MUKHERJEE and D. CARROLL, 2013 Donor DNA utilization during gene targeting with zinc-finger nucleases. G3: Genes| Genomes| Genetics **3:** 657-664.

BHAT, K. M., 1999 Segment polarity genes in neuroblast formation and identity specification during *Drosophila* neurogenesis. Bioessays **21:** 472-485.

BIBIKOVA, M., K. BEUMER, J. K. TRAUTMAN and D. CARROLL, 2003 Enhancing gene targeting with designed zinc finger nucleases. Science **300:** 764.

BIBIKOVA, M., M. GOLIC, K. G. GOLIC and D. CARROLL, 2002 Targeted chromosomal cleavage and mutagenesis in *Drosophila* using zinc-finger nucleases. Genetics **161:** 1169-1175.

BIEHS, B., V. FRANCOIS and E. BIER, 1996 The *Drosophila* short gastrulation gene prevents Dpp from autoactivating and suppressing neurogenesis in the neuroectoderm. Genes & development **10:** 2922-2934.

BITINAITE, J., D. A. WAH, A. K. AGGARWAL and I. SCHILDKRAUT, 1998 FokI dimerization is required for DNA cleavage. Proceedings of the National Academy of Sciences **95:** 10570-10575.

BOCH, J., H. SCHOLZE, S. SCHORNACK, A. LANDGRAF, S. HAHN *et al.*, 2009 Breaking the code of DNA binding specificity of TAL-type III effectors. Science **326:** 1509-1512.

BOONE, J. Q., and C. Q. DOE, 2008 Identification of *Drosophila* type II neuroblast lineages containing transit amplifying ganglion mother cells. Developmental neurobiology **68:** 1185-1195.

BOSSING, T., G. UDOLPH, C. Q. DOE and G. M. TECHNAU, 1996 The Embryonic Central Nervous System Lineages of *Drosophila melanogaster*: I. Neuroblast Lineages Derived from the Ventral Half of the Neuroectoderm. Developmental biology **179:** 41-64.

BOWMAN, S. K., V. ROLLAND, J. BETSCHINGER, K. A. KINSEY, G. EMERY *et al.*, 2008 The Tumor Suppressors Brat and Numb Regulate Transit-Amplifying Neuroblast Lineages in *Drosophila.* Developmental cell **14:** 535-546.

BRAND, A. H., and N. PERRIMON, 1993 Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. Development **118:** 401-415.

BROADUS, J., J. B. SKEATH, E. P. SPANA, T. BOSSING, G. TECHNAU *et al.*, 1995 New neuroblast markers and the origin of the aCC/pCC neurons in the *Drosophila* central nervous system. Mechanisms of development **53:** 393-402.

BUESCHER, M., S. L. YEO, G. UDOLPH, M. ZAVORTINK, X. YANG *et al.*, 1998 Binary sibling neuronal cell fate decisions in the *Drosophila* embryonic central nervous system are nonstochastic and require inscuteable-mediated asymmetry of ganglion mother cells. Genes & development **12:** 1858-1870.

BULLOCK, T., and G. A. HORRIDGE, 1965 Structure and function in the nervous systems of invertebrates.

BYRNE, S. M., L. ORTIZ, P. MALI, J. AACH and G. M. CHURCH, 2014 Multi-kilobase homozygous targeted gene replacement in human induced pluripotent stem cells. Nucleic acids research**:** gku1246.

CAMPOS-ORTEGA, J. A., 1995 Genetic mechanisms of early neurogenesis in *Drosophila melanogaster*. Molecular neurobiology **10:** 75-89.

CAPECCHI, M. R., 2005 Gene targeting in mice: functional analysis of the mammalian genome for the twenty-first century. Nat Rev Genet **6:** 507-512.

CARROLL, D., J. J. MORTON, K. J. BEUMER and D. J. SEGAL, 2006 Design, construction and in vitro testing of zinc finger nucleases. Nature protocols **1:** 1329-1341.

CARROLL, S. B., 1995 Homeotic genes and the evolution of arthropods and chordates. Nature **376:** 479-485.

CHEN, C. H., H. HUANG, C. M. WARD, J. T. SU, L. V. SCHAEFFER *et al.*, 2007 A synthetic maternal-effect selfish genetic element drives population replacement in *Drosophila*. Science **316:** 597-600.

CHEN, D., and D. M. MCKEARIN, 2003 A discrete transcriptional silencer in the bam gene determines asymmetric division of the *Drosophila* germline stem cell. Development **130:** 1159-1170.

CONG, L., F. A. RAN, D. COX, S. LIN, R. BARRETTO *et al.*, 2013 Multiplex genome engineering using CRISPR/Cas systems. Science **339:** 819-823.

DA SILVA, S., and F. WANG, 2011 Retrograde neural circuit specification by target-derived neurotrophins and growth factors. Current opinion in neurobiology **21:** 61-67.

DIAO, F., and B. H. WHITE, 2012 A novel approach for directing transgene expression in *Drosophila*: T2A-Gal4 in-frame fusion. Genetics **190:** 1139-1144.

DOE, C. Q., 1992 Molecular markers for identified neuroblasts and ganglion mother cells in the *Drosophila* central nervous system. Development **116:** 855-863.

DUMSTREI, K., C. NASSIF, G. ABBOUD, A. ARYAI and V. HARTENSTEIN, 1998 EGFR signaling is required for the differentiation and maintenance of neural progenitors along the dorsal midline of the *Drosophila* embryonic head. Development **125:** 3417-3426.

ELLIS, M. C., E. M. O'NEILL and G. M. RUBIN, 1993 Expression of *Drosophila* glass protein and evidence for negative regulation of its activity in non-neuronal cells by another DNA-binding protein. Development **119:** 855-865.

FRANCO, S. J., C. GIL-SANZ, I. MARTINEZ-GARAY, A. ESPINOSA, S. R. HARKINS-PERRY *et al.*, 2012 Fate-restricted neural progenitors in the mammalian cerebral cortex. Science **337:** 746-749.

GONG, W. J., and K. G. GOLIC, 2003 Ends-out, or replacement, gene targeting in *Drosophila*. Proceedings of the National Academy of Sciences **100:** 2556-2561.

GOODMAN, C. S., and C. Q. DOE, 1993 Embryonic development of the *Drosophila* central nervous system. The development of *Drosophila melanogaster* **2:** 1131-1206.

GRATZ, S. J., A. M. CUMMINGS, J. N. NGUYEN, D. C. HAMM, L. K. DONOHUE *et al.*, 2013 Genome engineering of *Drosophila* with the CRISPR RNA-guided Cas9 nuclease. Genetics **194:** 1029-1035.

GRATZ, S. J., F. P. UKKEN, C. D. RUBINSTEIN, G. THIEDE, L. K. DONOHUE *et al.*, 2014 Highly specific and efficient CRISPR/Cas9-catalyzed homology-directed repair in *Drosophila*. Genetics **196:** 961-971.

GRIFFITH, L. C., 2012 Identifying behavioral circuits in *Drosophila melanogaster*: moving targets in a flying insect. Current opinion in neurobiology **22:** 609-614.

HAMMONDS, A. S., C. A. BRISTOW, W. W. FISHER, R. WEISZMANN, S. WU *et al.*, 2013 Spatial expression of transcription factors in *Drosophila* embryonic organ development. Genome Biol **14:** R140.

HARRISON, M. M., B. V. JENKINS, K. M. O'CONNOR-GILES and J. WILDONGER, 2014 A CRISPR view of development. Genes & development **28:** 1859-1872.

HARTENSTEIN, V., and J. A. CAMPOS-ORTEGA, 1984 Early neurogenesis in wild-type *Drosophila melanogaster*. Wilhelm Roux's archives of developmental biology **193:** 308-325.

HARTENSTEIN, V., and J. A. CAMPOS-ORTEGA, 1985 Fate-mapping in wild-type *Drosophila melanogaster*. Wilhelm Roux's archives of developmental biology **194:** 181-195.

HAY, B. A., T. WOLFF and G. M. RUBIN, 1994 Expression of baculovirus P35 prevents cell death in *Drosophila*. Development **120:** 2121-2129.

HERNANDEZ, G., F. VALAFAR and W. E. STUMPH, 2007 Insect small nuclear RNA gene promoters evolve rapidly yet retain conserved features involved in determining promoter activity and RNA polymerase specificity. Nucleic acids research **35:** 21-34.

HIDALGO, A., A. R. LEARTE, P. MCQUILTON, J. PENNACK and B. ZHU, 2005 Neurotrophic and gliatrophic contexts in *Drosophila*. Brain, behavior and evolution **68:** 173-180.

HIROMI, Y., and W. J. GEHRING, 1987 Regulation and function of the *Drosophila* segmentation gene fushi tarazu. Cell **50:** 963-974.

HIRTH, F., B. HARTMANN and H. REICHERT, 1998 Homeotic gene action in embryonic brain development of *Drosophila*. Development **125:** 1579-1589.

HIRTH, F., S. THERIANOS, T. LOOP, W. J. GEHRING, H. REICHERT *et al.*, 1995 Developmental defects in brain segmentation caused by mutations of the homeobox genes orthodenticle and empty spiracles in *Drosophila*. Neuron **15:** 769-778.

HORN, C., B. JAUNICH and E. A. WIMMER, 2000 Highly sensitive, fluorescent transformation marker for *Drosophila* transgenesis. Development genes and evolution **210:** 623-629.

HSU, P. D., E. S. LANDER and F. ZHANG, 2014 Development and applications of CRISPR-Cas9 for genome engineering. Cell **157:** 1262-1278.

HUANG, J., P. GHOSH, G. F. HATFULL and Y. HONG, 2011 Successive and Targeted DNA Integrations in the *Drosophila* Genome by Bxb1 and φC31 Integrases. Genetics **189:** 391-395.

HUANG, J., W. ZHOU, A. M. WATSON, Y.-N. JAN and Y. HONG, 2008 Efficient ends-out gene targeting in *Drosophila*. Genetics **180:** 703-707.

HWANG, W. Y., Y. FU, D. REYON, M. L. MAEDER, S. Q. TSAI *et al.*, 2013 Efficient genome editing in zebrafish using a CRISPR-Cas system. Nature biotechnology **31:** 227-229.

ISSHIKI, T., M. TAKEICHI and A. NOSE, 1997 The role of the msh homeobox gene during *Drosophila* neurogenesis: implication for the dorsoventral specification of the neuroectoderm. Development **124:** 3099-3109.

ITO, K., and Y. HOTTA, 1992 Proliferation pattern of postembryonic neuroblasts in the brain of *Drosophila melanogaster*. Developmental biology **149:** 134-148.

IZERGINA, N., J. BALMER, B. BELLO and H. REICHERT, 2009 Postembryonic development of transit amplifying neuroblast lineages in the *Drosophila* brain. Neural Dev **4:** 44.

JACOB, J., C. MAURANGE and A. P. GOULD, 2008 Temporal control of neuronal diversity: common regulatory principles in insects and vertebrates? Development **135:** 3481-3489.

JASIN, M., 1996 Genetic manipulation of genomes with rare-cutting endonucleases. Trends in Genetics **12:** 224-228.

JENETT, A., G. M. RUBIN, T.-T. NGO, D. SHEPHERD, C. MURPHY *et al.*, 2012 A GAL4-driver line resource for *Drosophila* neurobiology. Cell reports **2:** 991-1001.

JESSELL, T. M., 2000 Neuronal specification in the spinal cord: inductive signals and transcriptional codes. Nature Reviews Genetics **1:** 20-29.

JIANG, Y., and H. REICHERT, 2012 Programmed cell death in type II neuroblast lineages is required for central complex development in the *Drosophila* brain. Neural development **7:** 1-15.

JINEK, M., K. CHYLINSKI, I. FONFARA, M. HAUER, J. A. DOUDNA *et al.*, 2012 A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. Science **337:** 816-821.

KANIA, M., A. BONNER, J. DUFFY and J. GERGEN, 1990 The *Drosophila* segmentation gene runt encodes a novel nuclear regulatory protein that is also expressed in the developing nervous system. Genes & Development **4:** 1701-1713.

KAO, C.-F., and T. LEE, 2010 Birth time/order-dependent neuron type specification. Current opinion in neurobiology **20:** 14-21.

KARCAVICH, R., and C. Q. DOE, 2005 *Drosophila* neuroblast 7−3 cell lineage: A model system for studying programmed cell death, Notch/Numb signaling, and sequential specification of ganglion mother cell identity. Journal of Comparative Neurology **481:** 240-251.

KIM, H., and J.-S. KIM, 2014 A guide to genome engineering with programmable nucleases. Nature Reviews Genetics.

KIM, J.-S., H. J. LEE and D. CARROLL, 2010 Genome editing with modularly assembled zinc-finger nucleases. Nature methods **7:** 91-91.

KIM, J. H., S.-R. LEE, L.-H. LI, H.-J. PARK, J.-H. PARK *et al.*, 2011 High cleavage efficiency of a 2A peptide derived from porcine teschovirus-1 in human cell lines, zebrafish and mice. PloS one **6:** e18556.

KIM, Y.-G., J. CHA and S. CHANDRASEGARAN, 1996 Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. Proceedings of the National Academy of Sciences **93:** 1156-1160.

KONDO, S., and R. UEDA, 2013 Highly improved gene targeting by germline-specific cas9 expression in *Drosophila*. Genetics **195:** 715-721.

KUMAR, A., B. BELLO and H. REICHERT, 2009 Lineage-specific cell death in postembryonic brain development of *Drosophila*. Development **136:** 3433-3442.

KURUSU, M., T. NAGAO, U. WALLDORF, S. FLISTER, W. J. GEHRING *et al.*, 2000 Genetic control of development of the mushroom bodies, the associative learning centers in the *Drosophila* brain, by the eyeless, twin of eyeless, and Dachshund genes. Proceedings of the National Academy of Sciences **97:** 2140-2144.

LACIN, H., Y. ZHU, B. A. WILSON and J. B. SKEATH, 2009 dbx mediates neuronal specification and differentiation through cross-repressive, lineage-specific interactions with eve and hb9. Development **136:** 3257-3266.

LAI, S.-L., and T. LEE, 2006 Genetic mosaic with dual binary transcriptional systems in *Drosophila*. Nature neuroscience **9:** 703-709.

LEE, T., and L. LUO, 1999 Mosaic analysis with a repressible cell marker for studies of gene function in neuronal morphogenesis. Neuron **22:** 451-461.

LEISMANN, O., and C. F. LEHNER, 2003 *Drosophila* securin destruction involves a D-box and a KEN-box and promotes anaphase in parallel with Cyclin A degradation. Journal of cell science **116:** 2453-2460.

LIN, S., C.-F. KAO, H.-H. YU, Y. HUANG and T. LEE, 2012 Lineage analysis of *Drosophila* lateral antennal lobe neurons reveals notch-dependent binary temporal fate decisions. PLoS biology **10:** e1001425.

LIN, S., S.-L. LAI, H.-H. YU, T. CHIHARA, L. LUO *et al.*, 2010 Lineage-specific effects of Notch/Numb signaling in post-embryonic development of the *Drosophila* brain. Development **137:** 43-51.

LIN, S., and T. LEE, 2012 Generating neuronal diversity in the *Drosophila* central nervous system. Developmental Dynamics **241:** 57-68.

LIN, S., E. C. MARIN, C.-P. YANG, C.-F. KAO, B. A. APENTENG *et al.*, 2013 Extremes of lineage plasticity in the *Drosophila* brain. Current Biology **23:** 1908-1913.

LUAN, H., N. C. PEABODY, C. R. VINSON and B. H. WHITE, 2006 Refined spatial manipulation of neuronal function by combinatorial restriction of transgene expression. Neuron **52:** 425-436.

LUI, J. H., D. V. HANSEN and A. R. KRIEGSTEIN, 2011 Development and evolution of the human neocortex. Cell **146:** 18-36.

LUNDELL, M. J., H.-K. LEE, E. PÉREZ and L. CHADWELL, 2003 The regulation of apoptosis by Numb/Notch signaling in the serotonin lineage of *Drosophila*. Development **130:** 4109-4121.

LUO, L., Y. J. LIAO, L. Y. JAN and Y. N. JAN, 1994 Distinct morphogenetic functions of similar small GTPases: *Drosophila* Drac1 is involved in axonal outgrowth and myoblast fusion. Genes & Development **8:** 1787-1802.

MALI, P., L. YANG, K. M. ESVELT, J. AACH, M. GUELL *et al.*, 2013 RNA-guided human genome engineering via Cas9. Science **339:** 823-826.

MILLER, J. C., S. TAN, G. QIAO, K. A. BARLOW, J. WANG *et al.*, 2011 A TALE nuclease architecture for efficient genome editing. Nature biotechnology **29:** 143-148.

MOLYNEAUX, B. J., P. ARLOTTA, J. R. MENEZES and J. D. MACKLIS, 2007 Neuronal subtype specification in the cerebral cortex. Nature reviews neuroscience **8:** 427-437.

MOSCOU, M. J., and A. J. BOGDANOVE, 2009 A simple cipher governs DNA recognition by TAL effectors. Science **326:** 1501-1501.

MUSSOLINO, C., J. ALZUBI, E. J. FINE, R. MORBITZER, T. J. CRADICK *et al.*, 2014 TALENs facilitate targeted genome editing in human cells with high specificity and low cytotoxicity. Nucleic acids research **42:** 6762-6773.

NASSIF, C., A. NOVEEN and V. HARTENSTEIN, 1998 Embryonic development of the *Drosophila* brain. I. Pattern of pioneer tracts. Journal of Comparative Neurology **402:** 10-31.

NERN, A., B. D. PFEIFFER, K. SVOBODA and G. M. RUBIN, 2011 Multiple new site-specific recombinases for use in manipulating animal genomes. Proceedings of the National Academy of Sciences **108:** 14198-14203.

NOVEEN, A., A. DANIEL and V. HARTENSTEIN, 2000 Early development of the *Drosophila* mushroom body: the roles of eyeless and dachshund. Development **127:** 3475-3488.

PANI, A. M., E. E. MULLARKEY, J. ARONOWICZ, S. ASSIMACOPOULOS, E. A. GROVE *et al.*, 2012 Ancient deuterostome origins of vertebrate brain signalling centres. Nature **483:** 289-294.

PEREANU, W., and V. HARTENSTEIN, 2006 Neural lineages of the *Drosophila* brain: a three-dimensional digital atlas of the pattern of lineage location and projection at the late larval stage. The Journal of neuroscience **26:** 5534-5553.

PFEIFFER, B. D., A. JENETT, A. S. HAMMONDS, T.-T. B. NGO, S. MISRA *et al.*, 2008 Tools for neuroanatomy and neurogenetics in *Drosophila*. Proceedings of the National Academy of Sciences **105:** 9715-9720.

PFEIFFER, B. D., T.-T. B. NGO, K. L. HIBBARD, C. MURPHY, A. JENETT *et al.*, 2010 Refinement of tools for targeted gene expression in *Drosophila*. Genetics **186:** 735-755.

PFEIFFER, B. D., J. W. TRUMAN and G. M. RUBIN, 2012 Using translational enhancers to increase transgene expression in *Drosophila*. Proceedings of the National Academy of Sciences **109:** 6626-6631.

PIERFELICE, T., L. ALBERI and N. GAIANO, 2011 Notch in the vertebrate nervous system: an old dog with new tricks. Neuron **69:** 840-855.

PORT, F., H. M. CHEN, T. LEE and S. L. BULLOCK, 2014 Optimized CRISPR/Cas tools for efficient germline and somatic genome engineering in *Drosophila*. Proc Natl Acad Sci U S A **111:** E2967-2976.

Ran, F. A., P. D. Hsu, C.-Y. Lin, J. S. Gootenberg, S. Konermann et al., 2013 Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. Cell 154: 1380-1389.

REIN, K., M. ZÖCKLER, M. T. MADER, C. GRÜBEL and M. HEISENBERG, 2002 The *Drosophila* standard brain. Current Biology **12:** 227-231.

REMPEL, J., 1975 evolution of the insect head: the endless dispute. Quaestiones Entomologicae**:** 7-25.

REN, X., J. SUN, B. E. HOUSDEN, Y. HU, C. ROESEL *et al.*, 2013 Optimized gene editing technology for *Drosophila melanogaster* using germ line-specific Cas9. Proceedings of the National Academy of Sciences **110:** 19012-19017.

ROGULJA-ORTMANN, A., and G. M. TECHNAU, 2008 Multiple roles for Hox genes in segment-specific shaping of CNS lineages. Fly **2:** 316-319.

RONG, Y. S., and K. G. GOLIC, 2000 Gene targeting by homologous recombination in *Drosophila*. Science **288:** 2013-2018.

RØRTH, P., 1998 Gal4 in the *Drosophila* female germline. Mechanisms of development **78:** 113-118.

ROUET, P., F. SMIH and M. JASIN, 1994 Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. Molecular and cellular biology **14:** 8096-8106.

SAUER, B., 1993 Manipulation of transgenes by site-specific recombination: Use of cre recombinase. Methods in Enzymology 225: 890-900.

SCHLAKE, T., and J. BODE, 1994 Use of mutated FLP recognition target (FRT) sites for the exchange of expression cassettes at defined chromosomal loci. Biochemistry **33:** 12746-12751.

SCHMID, A., A. CHIBA and C. Q. DOE, 1999 Clonal analysis of *Drosophila* embryonic neuroblasts: neural cell types, axon projections and muscle targets. Development **126:** 4653-4689.

SCHMIDT, H., C. RICKERT, T. BOSSING, O. VEF, J. URBAN *et al.*, 1997 The Embryonic Central Nervous System Lineages of *Drosophila melanogaster*. Developmental biology **189:** 186-204.

SCHMIDT-OTT, U., M. GONZÁLEZ-GAITÁN, H. JÄCKLE and G. M. TECHNAU, 1994 Number, identity, and sequence of the *Drosophila* head segments as revealed by neural elements and their deletion patterns in mutants. Proceedings of the National Academy of Sciences **91:** 8363-8367.

SCHMIDT-OTT, U., and G. M. TECHNAU, 1992 Expression of en and wg in the embryonic head and brain of *Drosophila* indicates a refolded band of seven segment remnants. Development **116:** 111-125.

SEBO, Z. L., H. B. LEE, Y. PENG and Y. GUO, 2013 A simplified and efficient germline-specific CRISPR/Cas9 system for *Drosophila* genomic engineering. Fly **8:** 8-7.

SIEGAL, M. L., and D. L. HARTL, 1996 Transgene coplacement and high efficiency site-specific recombination with the Cre/loxP system in *Drosophila*. Genetics **144:** 715.

SKEATH, J. B., 1998 The *Drosophila* EGF receptor controls the formation and specification of neuroblasts along the dorsal-ventral axis of the *Drosophila* embryo. Development **125:** 3301-3312.

SKEATH, J. B., 1999 At the nexus between pattern formation and cell−type specification: the generation of individual neuroblast fates in the *Drosophila* embryonic central nervous system. Bioessays **21:** 922-931.

SKEATH, J. B., and C. Q. DOE, 1998 Sanpodo and Notch act in opposition to Numb to distinguish sibling neuron fates in the *Drosophila* CNS. Development **125:** 1857-1865.

SKEATH, J. B., and S. THOR, 2003 Genetic control of *Drosophila* nerve cord development. Current opinion in neurobiology **13:** 8-15.

SMITH, J., M. BIBIKOVA, F. G. WHITBY, A. REDDY, S. CHANDRASEGARAN *et al.*, 2000 Requirements for double-strand cleavage by chimeric restriction enzymes with zinc finger DNA-recognition domains. Nucleic acids research **28:** 3361-3369.

SPANA, E. P., and C. Q. DOE, 1996 Numb antagonizes Notch signaling to specify sibling neuron cell fates. Neuron **17:** 21-26.

SPANA, E. P., C. KOPCZYNSKI, C. S. GOODMAN and C. Q. DOE, 1995 Asymmetric localization of numb autonomously determines sibling neuron identity in the *Drosophila* CNS. Development **121:** 3489-3494.

SPRADLING, A. C., 1993 Developmental genetics of oogenesis. The development of *Drosophila melanogaster* **1:** 1-70.

ST JOHNSTON, D., and C. NÜSSLEIN-VOLHARD, 1992 The origin of pattern and polarity in the *Drosophila* embryo. Cell **68:** 201-219.

SZYMCZAK, A. L., C. J. WORKMAN, Y. WANG, K. M. VIGNALI, S. DILIOGLOU *et al.*, 2004 Correction of multi-gene deficiency in vivo using a single 'self-cleaving' 2A peptide–based retroviral vector. Nature biotechnology **22:** 589-594.

TECHNAU, G. M., C. BERGER and R. URBACH, 2006 Generation of cell diversity and segmental pattern in the embryonic central nervous system of *Drosophila*. Dev Dyn **235:** 861-869.

TECHNAU, G. M., and J. A. CAMPOS-ORTEGA, 1985 Fate-mapping in wild-type *Drosophila melanogaster*. Wilhelm Roux's archives of developmental biology **194:** 196-212.

THOMAS, K. R., and M. R. CAPECCHI, 1987 Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. Cell **51:** 503-512.

TRUMAN, J. W., and M. BATE, 1988 Spatial and temporal patterns of neurogenesis in the central nervous system of *Drosophila melanogaster*. Developmental biology **125:** 145-157.

TRUMAN, J. W., W. MOATS, J. ALTMAN, E. C. MARIN and D. W. WILLIAMS, 2010 Role of Notch signaling in establishing the hemilineages of secondary neurons in *Drosophila melanogaster*. Development **137:** 53-61.

UDOLPH, G., K. LUER, T. BOSSING and G. M. TECHNAU, 1995 Commitment of CNS progenitors along the dorsoventral axis of *Drosophila* neuroectoderm. Science **269:** 1278-1281.

URBACH, R., 2007 A procephalic territory in *Drosophila* exhibiting similarities and dissimilarities compared to the vertebrate midbrain/hindbrain boundary region. Neural Dev **2:** 23.

URBACH, R., R. SCHNABEL and G. M. TECHNAU, 2003 The pattern of neuroblast formation, mitotic domains and proneural gene expression during early brain development in *Drosophila*. Development **130:** 3589-3606.

URBACH, R., and G. M. TECHNAU, 2003a Molecular markers for identified neuroblasts in the developing brain of *Drosophila*. Development **130:** 3621-3637.

URBACH, R., and G. M. TECHNAU, 2003b Segment polarity and DV patterning gene expression reveals segmental organization of the *Drosophila* brain. Development **130:** 3607-3620.

URBACH, R., and G. M. TECHNAU, 2004 Neuroblast formation and patterning during early brain development in *Drosophila*. Bioessays **26:** 739-751.

URBACH, R., D. VOLLAND, J. SEIBERT and G. M. TECHNAU, 2006 Segment-specific requirements for dorsoventral patterning genes during early brain development in *Drosophila*. Development **133:** 4315-4330.

VENKEN, K. J., and H. J. BELLEN, 2007 Transgenesis upgrades for Drosophila melanogaster. Development **134:** 3571-3584.

VENKEN, K. J., J. H. SIMPSON and H. J. BELLEN, 2011 Genetic manipulation of genes and cells in the nervous system of the fruit fly. Neuron **72:** 202-230.

VIKTORIN, G., N. RIEBLI and H. REICHERT, 2013 A multipotent transit-amplifying neuroblast lineage in the central brain gives rise to optic lobe glial cells in *Drosophila*. Developmental biology **379:** 182-194.

VON OHLEN, T., and C. Q. DOE, 2000 Convergence of dorsal, dpp, and egfr signaling pathways subdivides the drosophila neuroectoderm into three dorsal-ventral columns. Developmental biology **224:** 362-372.

WAKIYAMA, M., T. MATSUMOTO and S. YOKOYAMA, 2005 *Drosophila* U6 promoter-driven short hairpin RNAs effectively induce RNA interference in Schneider 2 cells. Biochemical and biophysical research communications **331:** 1163-1170.

WANG, Y.-C., J. S. YANG, R. JOHNSTON, Q. REN, Y.-J. LEE *et al.*, 2014 *Drosophila* intermediate neural progenitors produce lineage-dependent related series of diverse neurons. Development **141:** 253-258.

WONDERS, C. P., and S. A. ANDERSON, 2006 The origin and specification of cortical interneurons. Nature Reviews Neuroscience **7:** 687-696.

XUE, Z., M. REN, M. WU, J. DAI, Y. S. RONG *et al.*, 2014 Efficient Gene Knock-out and Knock-in with Transgenic Cas9 in *Drosophila*. G3: Genes| Genomes| Genetics **4:** 925-929.

YOUNOSSI-HARTENSTEIN, A., C. NASSIF, P. GREEN and V. HARTENSTEIN, 1996 Early neurogenesis of the *Drosophila* brain. Journal of Comparative Neurology **370:** 313-329.

YU, H.-H., C.-H. CHEN, L. SHI, Y. HUANG and T. LEE, 2009 Twin-spot MARCM to reveal the developmental origin and identity of neurons. Nature neuroscience **12:** 947-953.

YU, H. H., T. AWASAKI, M. D. SCHROEDER, F. LONG, J. S. YANG *et al.*, 2013a Clonal development and organization of the adult *Drosophila* central brain. Curr Biol **23:** 633-643.

YU, Z., H. CHEN, J. LIU, H. ZHANG, Y. YAN *et al.*, 2014 Various applications of TALEN-and CRISPR/Cas9-mediated homologous recombination to modify the *Drosophila* genome. Biology open**:** BIO20147682.

YU, Z., M. REN, Z. WANG, B. ZHANG, Y. S. RONG *et al.*, 2013b Highly efficient genome modifications mediated by CRISPR/Cas9 in *Drosophila*. Genetics **195:** 289-291.

ZHAO, C., W. DENG and F. H. GAGE, 2008 Mechanisms and functional implications of adult neurogenesis. Cell **132:** 645-660.

ZINYK, D. L., E. H. MERCER, E. HARRIS, D. J. ANDERSON, and A. L. JOYNER, 1998 Fate mapping of the mouse midbrain–hindbrain constriction using a site-specific recombination system. Current biology 8: 665-672.