

PXGS: a Poly-Transgene Expression System based on Mutually Exclusive Splicing of *Dscam*

Renee Yin Yu^{1,†}, Alyeri Bucio-Méndez^{1,†}, Brian E. Chen^{1,2,*}

¹ Centre for Research in Neuroscience, Research Institute of the McGill University Health Centre, Montréal, Québec, Canada, H3G 1A4.

² Departments of Medicine and Neurology & Neurosurgery, McGill University, Montréal, Québec, Canada, H3G 1A4.

† These authors contributed equally to this work.

* Corresponding author: Brian E. Chen, Montreal General Hospital, 1650 Cedar Ave, L7-224, Montréal, Québec, Canada, H3G 1A4, tel: 514-934-1934 x42379, fax: 514-934-8216, brian.chen@mcgill.ca

Running title: A poly-transgene expression system

Abstract

Biologists often need to investigate multiple genes simultaneously in an organism. However, it is currently not possible to express more than a few transgenes in an animal under conditional control. Here, we developed a technique based on the mutually exclusive splicing of the *Down Syndrome Cell Adhesion Molecule1* (*Dscam1*) gene in *Drosophila melanogaster* to achieve simultaneous transgene expression of 12 genes at a time. We show that the hypervariable *Dscam1* exon 4 region maintains its alternative splicing when placed in a UAS expression vector. Each of the twelve exon 4 alternates can be replaced with an exogenous gene of at least 10 kilobases and will express properly *in vivo* all under conditional genetic control. We demonstrate the expression of four different fluorophores placed in different exon 4 alternate positions in neural and non-neural cells *in vivo*. We validated the technique by rewiring *Drosophila* sensory neuron axons *in vivo* by simultaneously expressing several cell surface receptors within the neuron. This technology will also enable *Drosophila melanogaster* as a model system for synthetic biology research.

Significance Statement

It is not possible to express more than a few transgenes at a time in any multicellular organism. Our manuscript describes a technique to express more than 10 transgenes at a time under conditional control *in vivo* in *Drosophila*. This can have applications in neuroscience, synthetic biology, and biomanufacturing.

Introduction

Proteins operate and co-evolve together with other proteins in complexes. In structural biology the detection of a protein's structure within its multiprotein complex is important for understanding its biological function and evolutionary relationship (Ban et al., 2000; Bhattacharya, 2009; Stahl et al., 2024; Wiederstein et al., 2014). In cellular and systems biology, multiple molecules often need to be simultaneously manipulated to examine their roles and interactions within a cell. Multiple genes can be expressed in bacteria or yeast using a promoter for each gene, but repeats of the same promoter can result in homologous recombination (Agmon et al., 2009). In multi-cellular organisms such as animals, it can be challenging to express multiple transgenes.

Multiple genes can be expressed using a single polycistronic mRNA strand (Kozak, 1999). For example, an internal ribosome entry site (IRES) is used by the poliovirus to initiate translation (Pelletier and Sonenberg, 1988), but can be used for bi-cistronic expression of two genes. An IRES sequence can be placed between an upstream gene and a downstream gene so that a ribosome can bind not only to the 5' cap of the mRNA to translate the upstream gene, but also the IRES site to translate the downstream gene. However, the gene downstream of the IRES site is often translated at a much lower rate than the upstream gene and only two genes at a time is possible (Mizuguchi et al., 2000). Polycistronic expression can also be achieved using RNA sequences that encode cis-acting hydrolase element (CHYSEL) peptides (Doronina et al., 2008). CHYSEL polypeptides are commonly known as "2A" and "2A-like" peptides, and are used by RNA viruses to express each of its viral genes in a single mRNA strand. Synthesis of the CHYSEL polypeptide causes steric hindrance at the exit tunnel of the ribosome, causing the ribosome to skip the last bond of the CHYSEL peptide chain and restart on the next peptide (Donnelly et al., 1997; Martin D. Ryan GL, 2002; Ryan et al., 1991). Genes placed upstream and downstream of specific CHYSEL peptide sequences can be co-translationally co-expressed at an almost stoichiometric ratio (Lo et al., 2015). Up to four genes can be simultaneously expressed (quad-cistronic) using this technique, but at decreasing efficiency for each downstream gene (Liu et al., 2017). Finally, the MultiLabel technique in mammalian cells can express up to five genes (Kriz et al., 2010), and is based on the MultiBAC baculovirus expression system used in moth *Spodoptera frugiperda* cell lines (Bieniossek et al., 2012; Sari et al., 2016).

Alternative splicing is nature's way of producing molecular diversity from a single gene. An extreme example of alternative splicing is the *Drosophila Down syndrome cell adhesion molecule1 (Dscam1)* gene. Dscam is a cell surface receptor that is involved in several aspects of neural development, neural plasticity, and immune recognition (Schmucker and Chen, 2009). The *Dscam1* gene can produce tens of thousands of different isoforms through mutually exclusive splicing. It has 20 constitutive exons and 4 variable exon clusters: 4, 6, 9, and 17 which can be alternatively spliced in a mutually exclusive manner within each cluster (Schmucker et al., 2000) (**Figure 1a**). Within each of these variable exon clusters, the cell will select one and

only one of the possible alternatives to splice into the mRNA during each transcription event. In variable exon 4, there will be only 1 out of 12 alternates spliced, in variable exon 6 only 1 out of 48 will be chosen, in exon 9 only 1 out of 33 will be chosen, and in exon 17 one or the other alternate is chosen, all incorporated into a single mRNA strand (**Figure 1a**). This splicing occurs for every round of *Dscam1* transcription, so there will be many mRNA molecules that each contain a different exon variant at each of the variable exon clusters.

Here, we show that placing the entire *Dscam* exon 4 alternatively spliced cassette section into a DNA plasmid under UAS control, and replacing each of the twelve exon 4 variants with a gene, allows each gene to be expressed simultaneously in a cell. Cell- or tissue-specific expression is controlled using the Gal4/UAS system (Brand and Perrimon, 1993). We also show that alternative splicing of polycistronic dsRNA can induce RNAi *in vivo*. We validate our technology by rewiring *Drosophila* sensory neurons axons *in vivo* by simultaneously expressing several large cell surface receptors solely within the neuron. We call this new technology the poly-transgene expression system (PXGS).

Results

Dscam mutually exclusive splicing is conserved in the PXGS system

We first sought to verify that *Dscam* variable exon 4 would be properly spliced within an exogenous DNA construct. We extracted the last 300 bases of *Dscam* exon 3 through variable exon 4 to the first 30 bases of exon 5 from *Drosophila melanogaster* genomic DNA. A 6× *Histidine* DNA sequence was added to the end of the exon 5 sequence as a genetic tag. We also wanted to test whether *Dscam* alternative splicing would be preserved after transcription under a UAS promoter (**Figure 1b**), to allow it to be expressed conditionally using a Gal4 driver (Brand and Perrimon, 1993). We inserted the 6.7 kilobasepair *Dscam* exon 4 fragment into the expression vector *pJFRC7-20XUAS*, and co-transfected it along with *Actin5C-Gal4* into *Drosophila melanogaster* S2 cells (**Figure 1c**). To determine whether the endogenous exon 4 alternates were properly spliced, we performed reverse transcriptase PCR (RT-PCR) using the 6× *Histidine* genetic tag to distinguish the UAS-transcripts from the endogenous *Dscam* mRNA transcripts. We found that in S2 cells transfected with *UAS-DscamExon4*, the exogenous Exon 4 variants were produced at the expected sizes as the endogenous Exon 4 alternates (**Figure 1c**).

Inserting fluorescent protein genes in the endogenous exon 4 alternates

The mutually exclusive splicing of *Dscam* exon 4 is determined by its intronic structures (Hong et al., 2021; Krehling and Graveley, 2005; Xu et al., 2019; Yue et al., 2016), but it is not certain that the exonic 4 sequences do not contribute. We first sought to insert the *green fluorescent protein* (*GFP*) transgene inside of (i.e., surrounded by) the endogenous exon 4 alternates, which would result in the endogenous sequences being treated as untranslated regions (UTRs) of mRNA flanking the transgene. Our goal was to then systematically narrow down the

minimum endogenous *Dscam* exon 4 sequences required for proper splicing. In parallel, we tried to completely replace the entire exon 4.1 alternate with the *GFP* gene. To our surprise, the *GFP* mRNA and functional fluorescence were expressed in S2 cells when the entire exon 4.1 was replaced (**Supplemental Figure 1**). On any given round of transcription of the *UAS-DscamExon4_GFP^{4.1}* vector, the probability that the *GFP* gene would be expressed is 1/12 (8% chance), compared to a 92% chance that any of the other 11 exon 4 alternates would be spliced (but not translated). Indeed, it would take 34 rounds of transcription to have a 95% chance that the *GFP* gene would be spliced at least once. Nevertheless, we show that only the intronic sequences within variable exon 4 are required for proper mutually exclusive splicing.

Replacing exon 4 variants with fluorophores

We next sought to verify that other exon 4 variants could be completely replaced to express any gene of interest within our *UAS-DscamExon4* construct. We replaced alternate exons 4.1, 4.2, 4.11, and 4.12 with *GFP*, *blue fluorescent protein* with a nucleolar localization signal (*BFP_{nols}*), *near-infrared fluorescent protein* (*iRFP*), and *red fluorescent protein* (*RFP*) transgenes, respectively. After verification of their expression in S2 cells using RT-PCR, we generated a transgenic fly containing the *UAS-DscamExon4* construct, *UAS-DscamExon4_GFP^{4.1}-BFP_{nols}^{4.2}-iRFP^{4.11}-RFP^{4.12}*. We crossed this fly to the pan-neuronal Gal4 driver *nSyb-Gal4*, and found that all four fluorophores were weakly to moderately expressed but not in all neurons (**Figure 2**). If Exon 4 alternative splicing is random, then after multiple rounds of transcription eventually all four colors should be expressed. If Exon 4 splicing is deterministic, then some neurons may never express the splicing factors that were to select the missing colors. In other words, some neurons may not express specific splicing factors to allow for expression of specific Exon 4 alternates. An important caveat for this *UAS-DscamExon4* construct is that for each round of transcription driven by the *nSyb-Gal4*, a neuron may splice any of the eight Exon 4 endogenous alternates (i.e., producing non-fluorescence) or any of the four fluorophores, so that the “signal” is diluted and divided, respectively, compared to if *nSyb-Gal4* were driving expression of a standard *UAS-fluorescent protein* construct.

Regardless if *Dscam* Exon 4 is random or deterministic, we show that Exon 4 alternates can be replaced by a gene of interest for poly-transgene expression and can be conditionally controlled by Gal4 expression. After transcription and splicing, any extraneous DNA surrounding the transgene such as Exon 3, Exon 5, and the 6× *Histidine* tag become untranslated mRNA regions (UTRs) and are not synthesized by the ribosome, as each transgene contains their own start and stop codons.

PXGS can drive expression in non-neuronal cells

Dscam is expressed in all cells, based on gene expression atlases for *Drosophila*, albeit at varying levels (Corrales et al., 2022; Li et al., 2022). *Dscam* has traditionally been studied in the nervous and immune systems. To verify that our poly-transgene expression system (PXGS) can

be used in cells beyond neurons and immune cells, we generated a multi-color PXGS reagent to label any cell that expresses *Dscam* by replacing ten exon 4 alternates with fluorophores. We differentially localized GFP, BFP, RFP, and iRFP to the membrane (mCD8::GFP), the nucleus (nols), or the mitochondria (COX8::RFP), and generated the construct *PXGS_iRFP^{nols}^{4.1}-COX8::RFP^{4.2}-BFP^{nols}^{4.3}-mCD8::GFP^{4.4}-iRFP^{nols}^{4.5}-COX8::RFP^{4.6}-mCD8::GFP^{4.8}-BFP^{nols}^{4.9}-iRFP^{nols}^{4.10}-BFP^{nols}^{4.12}*. We verified the PXGS expression in S2 cells by co-transfection with *Actin5C-Gal4* (**Supplemental Figure 2a**) and proceeded to generate transgenic flies. We crossed the *UAS-PXGS_iRFP^{nols}^{4.1}-COX8::RFP^{4.2}-BFP^{nols}^{4.3}-mCD8::GFP^{4.4}-iRFP^{nols}^{4.5}-COX8::RFP^{4.6}-mCD8::GFP^{4.8}-BFP^{nols}^{4.9}-iRFP^{nols}^{4.10}-BFP^{nols}^{4.12}* flies to the pan-neuronal *nSyb-Gal4* and imaged the brains to verify expression *in vivo* (**Supplemental Figure 2b**). Neurons were strongly fluorescent with green at the cell membrane, intracellular punctate red mitochondria, and overlapping blue and infrared nuclei. However, the blue channel was often dominated by autofluorescence of the tracheal tubes.

Crossing the *UAS-PXGS_iRFP^{nols}^{4.1}-COX8::RFP^{4.2}-BFP^{nols}^{4.3}-mCD8::GFP^{4.4}-iRFP^{nols}^{4.5}-COX8::RFP^{4.6}-mCD8::GFP^{4.8}-BFP^{nols}^{4.9}-iRFP^{nols}^{4.10}-BFP^{nols}^{4.12}* flies to the pan-glial *Repo-Gal4* driver revealed expression of all four fluorophores across the brain (**Figure 3**). We crossed these PXGS flies to *Tubulin-Gal4* for ubiquitous expression across the animal. Fluorescence expression was highest in the brain, but we also observed fluorescence in muscle. Multi-nucleated muscle cells were labelled with blue and near-infrared fluorescence, green cell membrane, and red mitochondria puncta (**Figure 3**). Similar to previous studies, we observed the mCD8::GFP expression concentrated at the immediate area surrounding the nucleus (Ralston and Hall, 1989).

Functional expression of PXGS transgenes

The genes for fluorescent proteins are ≤ 1 kilobase in size. To demonstrate that PXGS can express large, functionally relevant genes, we chose eight cell surface receptors to mis-express in a mechanosensory neuron. The eight genes were *Bsg*, *dpr8*, *dpr12*, *kek1*, *kirre*, *sli*, *Toll-6*, and *tutl*. We created three PXGS constructs: *UAS-PXGS_dpr8^{4.1}-dpr12^{4.2}-Gal4^{4.3}*, *UAS-PXGS_kek1^{4.7}-kirre^{4.8}-tutl^{4.9}*, and *UAS-PXGS_Toll-6^{4.10}-Bsg^{4.11}-sli^{4.12}* and verified their mRNA expression in S2 cells (**Supplemental Figure 3**) before generating fly lines. The gene for Gal4 itself was inserted into the Exon 4.3 position to create a transcriptional positive feedback loop for continuous gene expression. We used the *455-Gal4* driver line to mis-express these cell surface receptors solely within the pSc mechanosensory neuron (Cvetkovska et al., 2013; Dos Santos et al., 2019; Kays et al., 2014; Neufeld et al., 2011). The pSc mechanosensory neuron has a stereotyped axonal targeting pattern within the central nervous system, which can be used as an assay to identify molecules involved in axonal growth and targeting. However, the *Dscam* receptor and isoforms are required for proper neuronal wiring (Chen et al., 2006), so overexpression of the PXGS constructs may interfere with the alternative splicing of endogenous *Dscam* isoforms by competing for the same splice factors, in a “sponging” effect. To account for this, we crossed the *455-Gal4* driver to the *UAS-PXGS_iRFP^{nols}^{4.1}-COX8::RFP^{4.2}-BFP^{nols}^{4.3}-*

mCD8::GFP^{4.4}-iRFP_{nols}^{4.5}-COX8::RFP^{4.6}-mCD8::GFP^{4.8}-BFP_{nols}^{4.9}-iRFP_{nols}^{4.10}-BFP_{nols}^{4.12} flies (herein referred to as *PXGS_fluorophores*) as our control lines (**Figure 4a**). We found that these control *455-Gal4 / PXGS_fluorophores* flies had significantly shorter total branch lengths of the pSc axon ($p < 0.05$, ANOVA) compared to *455-Gal4 / +* controls ($n = 30$). However, this is likely due to genetic background differences, and not a “sponging” effect of PXGS on *Dscam* splicing factors, because the *PXGS_fluorophore / +* flies (with no Gal4) ($n = 15$) also had significantly shorter pSc branch lengths than *455-Gal4 / +* flies. The difference in axonal branch lengths between *PXGS_fluorophore / +* flies and *455-Gal4 / PXGS_fluorophores* flies was not significant (**Figure 4d**).

Compared to the control flies ($n = 13$), mis-expression of *dpr8* and *dpr12* ($n = 20$) or *Toll-6*, *Bsg*, and *sli* ($n = 15$) within the pSc resulted in a significant increase in axonal targeting errors ($p < 0.0001$, *t*-test) (**Figures 4b, c**). The axons frequently failed to extend posteriorly or anteriorly on the contralateral side. In extreme cases, the axon would stall at the entry point (**Figure 4b**). This resulted in significantly shorter total branch lengths and total branch numbers in these two fly lines, compared to *455-Gal4 / PXGS_fluorophores* control ($p < 0.05$, ANOVA) (**Figure 4d**). Mis-expression of *kek1*, *kirre*, and *tutl* simultaneously in the pSc neuron ($n = 15$) frequently resulted in the axon exiting the thoracic ganglion (**Figures 4b, c**). Thus, PXGS can express multiple, functionally relevant, receptor genes in a single cell.

Discussion

We created the poly-transgene expression system to allow any *Drosophila* biologist to simultaneously express multiple genes under conditional control. Compared to other multi-gene expression strategies, PXGS offers two distinct advantages. First, it can express many more transgenes than what was previously achievable. Each PXGS construct can express 12 transgenes, so that instead of a standard triple-transgenic animal containing three transgenes, a PXGS animal would have 36 transgenes. Second, the molecular biology to create a PXGS is straightforward and easier to work with. For example, using 2A peptide DNA sequences for more than two genes can be technically challenging due to the repetition of the 2A sequences, which can confound the primer design step, the plasmid assembly, and the sequence verification. This difficulty is eliminated in PXGS. The DNA sequences of the intronic regions spanning each of the 12 exon alternates are different enough that molecular assembly primers do not have multiple binding sites. The turnaround time to generate a PXGS construct is around 3–4 weeks.

Whether a specific cell expresses all 12 Exon 4 alternates for PXGS is dependent on three things. First, the expression level of the Gal4 driver dictates how many rounds of PXGS transcription will occur. This can be supplemented by including the gene for Gal4 within one of the alternates (**Figure 4a**). Second, if the cell expresses *Dscam*, then it should express Exon 4 splicing factors, but all of the splicing factors involved are not known. If *Dscam* exon 4 splicing is deterministic in that cell, then it may never splice all 12 Exon 4 alternates. However, cells that

do not express the correct splice factors could be made to, by expressing the correct splicing factors using PXGS itself. Obviously this is dependent on knowing all of the splice factors and mechanisms involved in Exon 4. Additionally, mis-expressing *Dscam* splicing factors would interfere with the endogenous *Dscam* splicing function and may adversely affect the cell given that those Exon 4 alternates were specifically not spliced. Third and similarly, if *Dscam* Exon 4 splicing is random, there still may be insufficient splicing factors despite a high transcription rate. A high transcription rate of PXGS may cause one Exon 4 alternate to be repeatedly selected due to the constitutive recruitment or high local concentration of its splicing factor(s).

For each round of transcription, the probability that any given Exon 4 alternate is not spliced is 11 out of 12. The number of endogenous *Dscam* isoforms expressed within a single cell can be used as a minimum number of rounds of transcription that must have occurred, and this is approximately 10 – 1000 isoforms (Schmucker and Chen, 2009). For ten rounds of transcription, the probability that a specific Exon 4 alternate is not spliced is $(11/12)^{10}$, or 42%. After 50 rounds of transcription, the probability that a specific exon 4 alternate is not spliced is $(11/12)^{50}$, or 1%.

Removing variants in Exon 4 does not affect splicing (Dong et al., 2023). Thus, PXGS can be customized to exactly the number of variants needed, by removing or duplicating the Exon 4 alternates, mimicking what has occurred through evolution (Brites et al., 2013; Lee et al., 2010). If *Dscam* Exon 4 splicing is random, then modifying the number of alternates in PXGS is less likely to negatively affect the cell.

For systems level biological investigations, loss-of-function experiments on multiple genes simultaneously may be required. RNAi is a common approach used for loss-of-function analyses (Qiao et al., 2018). UAS-based expression of dsRNA allows for RNAi-mediated knockdown in specific cells. We generated a *UAS-PXGS_dsRNA-RFP^{4.1}-BFP_{nols}^{4.2}-Gal4^{4.3}-Shibire^{4.6}-iRFP^{4.11}-RFP^{4.12}* fly to knockdown the RFP expression in position 4.12, while leaving the BFP_{nols} in position 4.2 unaffected. Crossing this fly to *nSyb-Gal4*, we found that RFP was not expressed in the brain, but nuclear BFP was clearly visible (**Supplemental Figure 4**). The strong BFP signal was likely due to the Gal4 in position 4.3 that created a positive feedback loop, but this may have also increased the off-target effects of the RNAi, as no iRFP was observed. These experiments demonstrate the potential to use PXGS for polycistronic gene knockdown. Thus, other RNAs may be properly spliced and processed within PXGS, such as guideRNAs, circular RNA, complex secondary structures, or other non-coding RNAs.

The PXGS system will also allow *Drosophila melanogaster* to be a new model system in synthetic biology (**Supplemental Figure 5**), and for biosynthesis and biomanufacturing in S2 cells (Coker et al., 2022). In addition, all animals in the Tetraconata (Pancrustacea) clade, which is nearly all arthropods, have a hypervariable *Dscam* gene with mutually exclusive splicing (Brites et al., 2013; Lee et al., 2010; Schmucker and Chen, 2009). The PXGS technology can be

adapted to any insect or crustacean to express multiple transgenes, such as in Sf9 cells from the *Spodoptera frugiperda* fall armyworm, or in High Five cells from the cabbage looper *Trichoplusia ni* for recombinant protein production *in vitro*, or *in vivo* in mosquitos, silkworm moth, honeybees, or aquaculture shrimp.

Methods

Inserting the Poly-transgene Expression System (PXGS) into plasmids

For simplicity, we refer to *Dscam1* as *Dscam* throughout, because the *Dscam1* gene is the only one of the four *Drosophila melanogaster* *Dscam* gene paralogs that has the hypervariable mutually exclusive alternative splicing. We amplified the entire *Dscam* exon 4 region along with 300 basepairs (bp) at the 3' end of *Dscam* exon 3 and 30 bp at the 5' start of exon 5 from *Drosophila melanogaster* genomic DNA, totaling 6.7 kilobasepairs. The PCR was carried out for 1 min at 94°C, 1 min at 55°C, and 15 min at 72°C for 35 cycles using CloneAmp (Takara Bio Inc. USA, Mountain View, CA) (Celotto and Graveley, 2001; Haussmann et al., 2019; Kreahtling and Graveley, 2005). There are several ATG codons (coding for Methionine) within exon 3, which would then create several unintended translation start sites, adding up to 70 unwanted, in-frame amino acids attached to the N-terminus of every PXGS transgene. To eliminate this problem, three point-mutations were inserted into exon 3 (**Supplementary Table 1**).

The *Dscam* fragment was then inserted into the pJFRC7-20XUAS-IVS-mCD8::GFP expression vector (Pfeiffer et al., 2010) using Gibson assembly (Gibson et al., 2009). The fragment was inserted downstream of the UAS sequences and replaced the mCD8::GFP sequence. A 6×Histidine tag was added at the end of exon 5 to distinguish PXGS sequences in reverse transcriptase PCR. Gibson assembly and transformation were performed with the In-Fusion® HD Cloning Kit and Stellar™ Competent Cells following the kit instructions (Takara Bio Inc. USA, Mountain View, CA).

Inserting transgenes into PXGS

Transgene insertions into the exon 4 alternates within our different PXGS plasmids were created using the automated Molecular Assembly feature in [GeneDig.org](https://www.genedig.org) (Suciu et al., 2015), primarily using gene synthesis, Gibson assembly, and restriction enzyme digest and ligation. Exon 4 variants were replaced with fluorophore genes codon optimized for *Drosophila*, or cell surface molecule genes, or hairpin RNA sequences.

The *GFP* gene was based on the *mNeonGreen* gene (Ceolin et al., 2020). The *BFP* and *RFP* genes were obtained from a plasmid previously generated in our lab (Lo and Chen, 2019). The *iRFP* gene was based on *miRFP670* (Shcherbakova et al., 2016) and codon optimized for *Drosophila* and synthesized by Integrated DNA Technology (Coralville, IA). To sequester the fluorophores into subcellular compartments to help distinguish their expression, we used transmembrane CD8 (mCD8) for membrane localization, a nucleolar localization signal (nols) for nuclear labelling, and COX8 for mitochondrial localization.

Sequences for the eight cell surface receptors, *Bsg* (798 bp), *dpr8* (891 bp), *dpr12* (1,035 bp), *kek1* (2,643 bp), *kirre* (2,871 bp), *sli* (4,443 bp), *Toll-6* (4,545 bp), and *tut1* (2,892 bp) were obtained from RT-PCR on wildtype flies using oligo d(T) reverse primer to generate the cDNA, and gene-specific forward and reverse primer to amplify the coding sequences. All DNA constructs were verified by colony PCR and Sanger sequencing. All other transgenes used in PXGS were based on their amino acid sequence and then codon optimized for *Drosophila* and synthesized by Integrated DNA Technology (Coralville, IA).

Cell culture

All PXGS transgenes were first verified using RT-PCR after transfection into S2 cells. *Drosophila* S2 cells were cultured at 25°C in Ex-Cell 420 Medium (Sigma-Aldrich, St. Louis, MO). Cells were transfected with 5 µg of plasmid DNA in 35 mm dishes using Lipofectamine (ThermoFisher, Waltham). 48 hours after transfection, RNA was extracted from the S2 cells and reverse transcription was performed on transfected and untransfected cells. PXGS-specific reverse primers for the 6×His tag or for the overlap between exon 5 and the vector were used to identify PXGS transgenes, and oligo-d(T) primers were used for endogenous *Dscam* mRNA alone. Endogenous *Dscam* exon 4 was used as the positive controls, and PXGS plasmid with no *Actin5C-Gal4* transfection was used as the negative controls. PCR was performed using gene-specific forward and reverse primers.

Fly stocks

To reduce variability due to environmental conditions and genetic background, flies were reared at 25°C on standard cornmeal under 12h light / 12h dark cycles. To express PXGS fluorophores in different cell types, the pan-neuronal *nSyb-Gal4*, glial *Repo-Gal4*, ubiquitous *Tubulin-Gal4*, and scutellar *455-Gal4* drivers were used. PXGS plasmids that were verified for expression in S2 cells were used to make transgenic animals. Transgenic *Drosophila* lines were generated using PhiC31 integrase-mediated transgenesis performed by BestGene, Inc (Chino Hills, CA).

Carbocyanine dye labeling and *Drosophila* imaging sample preparation

Drosophila mechanosensory axonal arbors were labelled using carbocyanine dye injection onto the pSc neurons, as previously described (Chen et al., 2006; Cvetkovska et al., 2013; Dos Santos et al., 2019; Kays et al., 2014; Neufeld et al., 2011). The left and right pSc bristles of 2-day old female flies were plucked and the thorax was left overnight in fixative solution (3.7% paraformaldehyde in 0.2 M carbonate-bicarbonate buffer, pH 9.5). The left and right pSc neurons were labelled using the carbocyanine tracers, DiI or DiD (ThermoFisher). The flies were left undisturbed and partially immersed in 0.2 M carbonate-bicarbonate buffer at pH 9.5. After 48 hours, the thoracic ganglion was dissected and imaged.

Image acquisition

Fluorescent images were acquired using an AxioScope A1 epifluorescence microscope (Carl Zeiss) with a 40× objective (N.A. 1.0) or an FV1000 laser-scanning confocal microscope (Olympus) with a 40× objective (N.A. 1.3). For mechanosensory axon imaging, 10–20 images were taken at different focal planes. Maximal intensity z-projections were stacked to generate the final image. Images were adjusted only for contrast and brightness when required. Transmitted light images were used to evaluate dissection and dye labeling quality and to measure thoracic ganglia width.

Image analysis

Only images without central nervous system damage or surface occlusions were included for data analysis. Measurements of axonal branch length and branch number were performed blind to genotype by randomly distributing the axonal arbor data of the total dataset. Custom-written software in MATLAB was used to perform the quantitative image analysis.

To determine the baseline variability in wildtype pSc mechanosensory neurons, the total number of branches and branch lengths per arbor were measured from thirty control 455-Gal4 / + animals. A skeleton of the control 455-Gal4 / + axonal arbor was then created based on the frequency of occurrence of each branch, which was found to be consistent with our previous measurements (Chen et al., 2006; Cvetkovska et al., 2013; Dos Santos et al., 2019; Kays et al., 2014; Neufeld et al., 2011). We defined an ectopic branch as any branch that has less than 30% occurrence in the wildtype. Representative examples of axonal arbors in figures were randomly chosen from the sets of all axonal arbor images of the genotype.

Statistical analysis

Statistical significance for total axonal branch length and number of branches was determined using one-way ANOVA followed by Dunnett's multiple comparisons test. Statistical significance for frequency of occurrence was determined using a two-tailed *t*-test for proportions. Statistical analysis was performed using GraphPad Prism version 8.0.0 for Windows (GraphPad Software, San Diego, California USA) and SPSS (version 25). The datasets, analyses, and materials used in the current study are available from the corresponding author upon request.

Acknowledgments

The authors thank Farida Emran for assistance with experiments. This work was supported by a grant from the Natural Sciences and Engineering Research Council of Canada (2020-04902, B.E.C.). The funding agency had no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Author Contributions

B.E.C. designed the experiments and supervised the project. R.Y.Y. and A.B-M. performed experiments and analyzed the data. R.Y.Y., A.B-M., and B.E.C. wrote the manuscript.

Competing Interest Statement

R.Y.Y., A.B-M., and B.E.C. are inventors on a pending patent describing the system and materials in this manuscript.

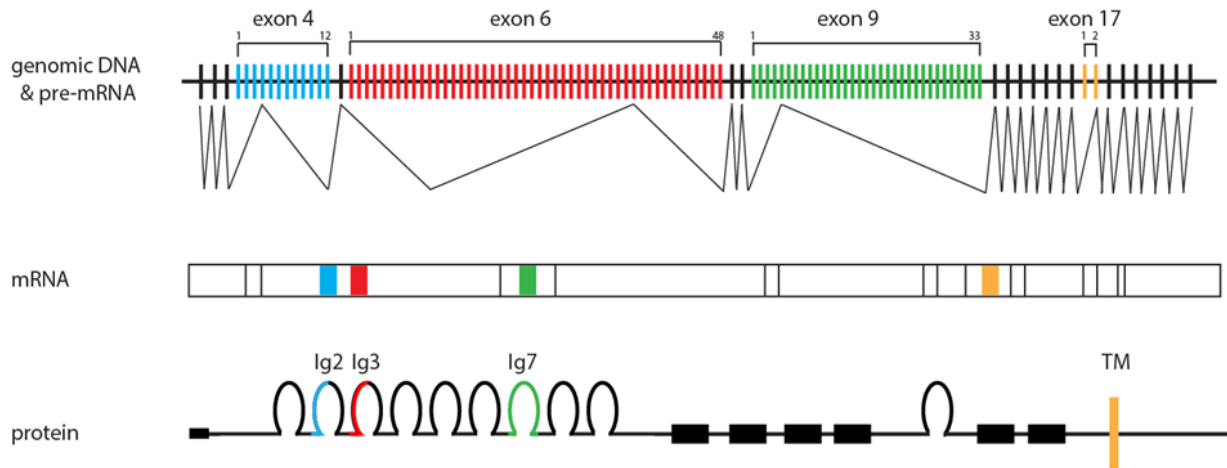
The data that support the findings of this study are available from the corresponding author upon request. Correspondence and requests for materials should be addressed to brian.chen@mcgill.ca.

Supplementary Information is provided as five supplementary figures and one supplementary table.

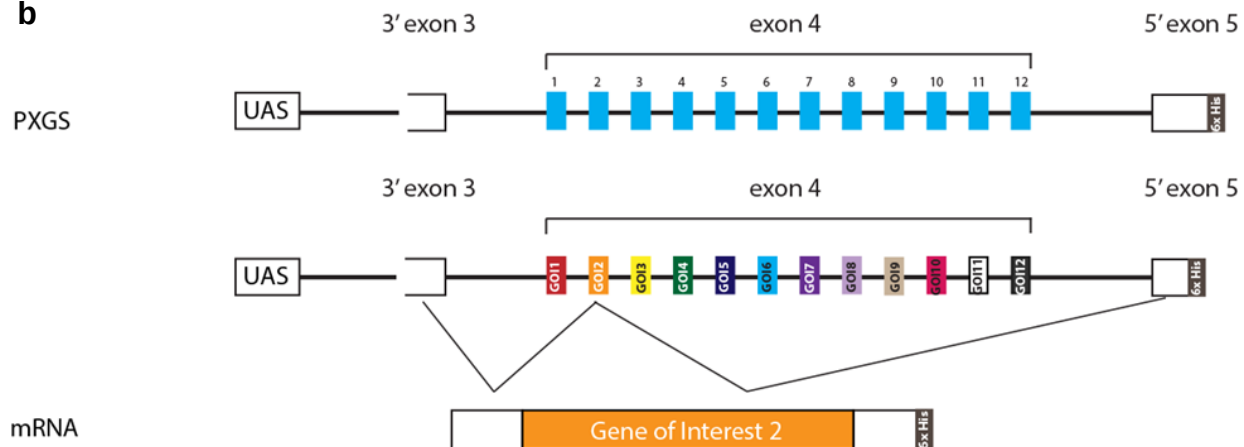
Figures

a

Drosophila melanogaster Dscam



b



c

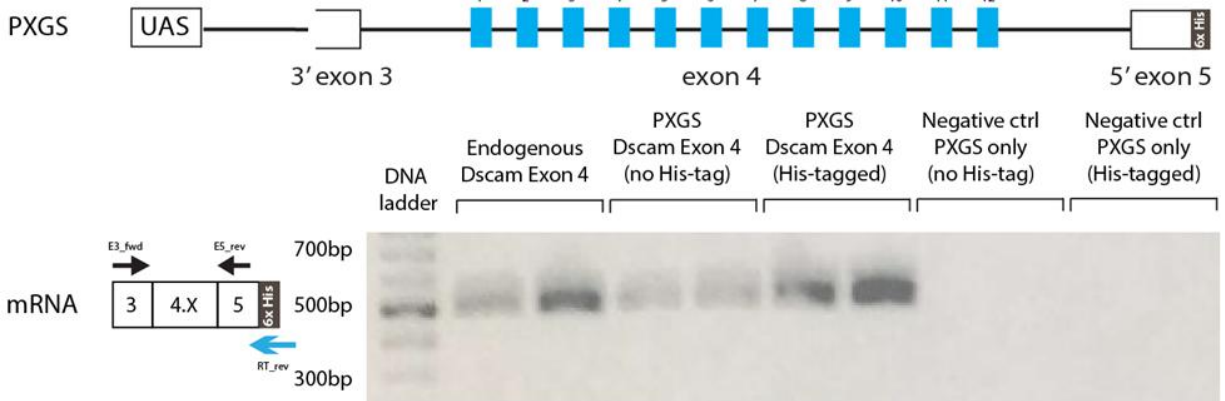


Figure 1, PXGS takes advantage of the mutually exclusive splicing property of *Dscam*.

a, Endogenous *Dscam1* mRNA is mutually exclusively spliced in four exons during transcription, exons 4, 6, 9, and 17. For each of the mutually exclusively spliced exons, with

every round of transcription, one and only one of the possible exon variants is in the final mRNA product and ultimately the protein product. **b**, The entire *Dscam* exon 4 cluster is inserted downstream of a *UAS* sequence to create the *PXGS* vector. A 6× histidine tag is added at the end of exon 5 to distinguish *PXGS* expression from endogenous *Dscam* exon 4 expression. Any gene of interest (GOI) can then be inserted to replace any of the 12 exon 4 alternates for expression. With each round of transcription, only a single transgene is selected for incorporation into the final mRNA product. **c**, Exon 4 mutually exclusive splicing is maintained in *PXGS*. *PXGS* vectors with and without a 6× histidine tag were co-transfected with *Actin5c*-Gal4 in *Drosophila* S2 cells. As a positive control, reverse transcription (RT) of Exon 4 was performed on untransfected S2 cells to identify the endogenous *Dscam* Exon 4 mRNA. RT using *PXGS*-specific reverse primers on S2 cells was performed, with the no *Actin5c*-Gal4 co-transfection (*PXGS* only) as negative controls. PCR of Exon 4 in the *PXGS* vectors with and without a 6×His-tag showed DNA bands at the expected 500 bp to 600 bp sizes.

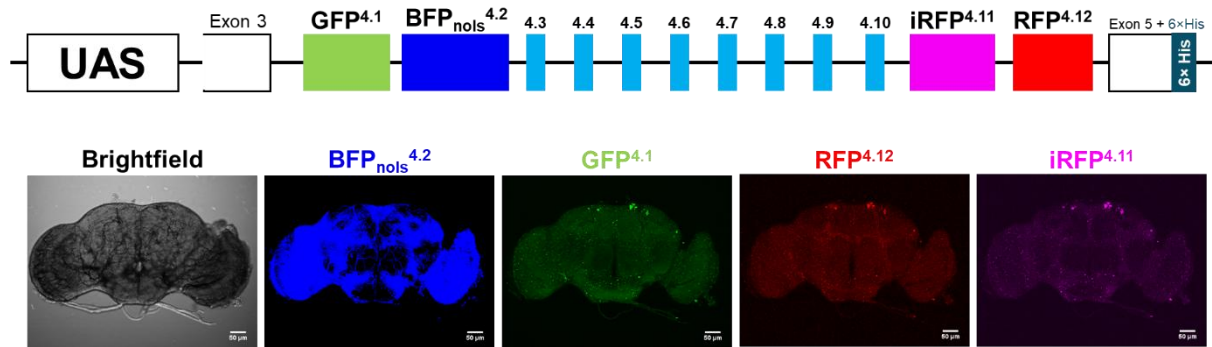


Figure 2, PXGS can express multiple fluorophores simultaneously *in vivo*.

Crossing the pan-neuronal *nSyb-Gal4* driver line with the transgenic fly *UAS-PXGS_GFP^{4.1}-BFP_{nols}^{4.2}-iRFP^{4.11}-RFP^{4.12}* demonstrates that nearly all neurons can splice and express multiple transgenes *in vivo*. However, the fluorescence intensity for any single color was low as its production was divided among 12 Exon 4 alternates.

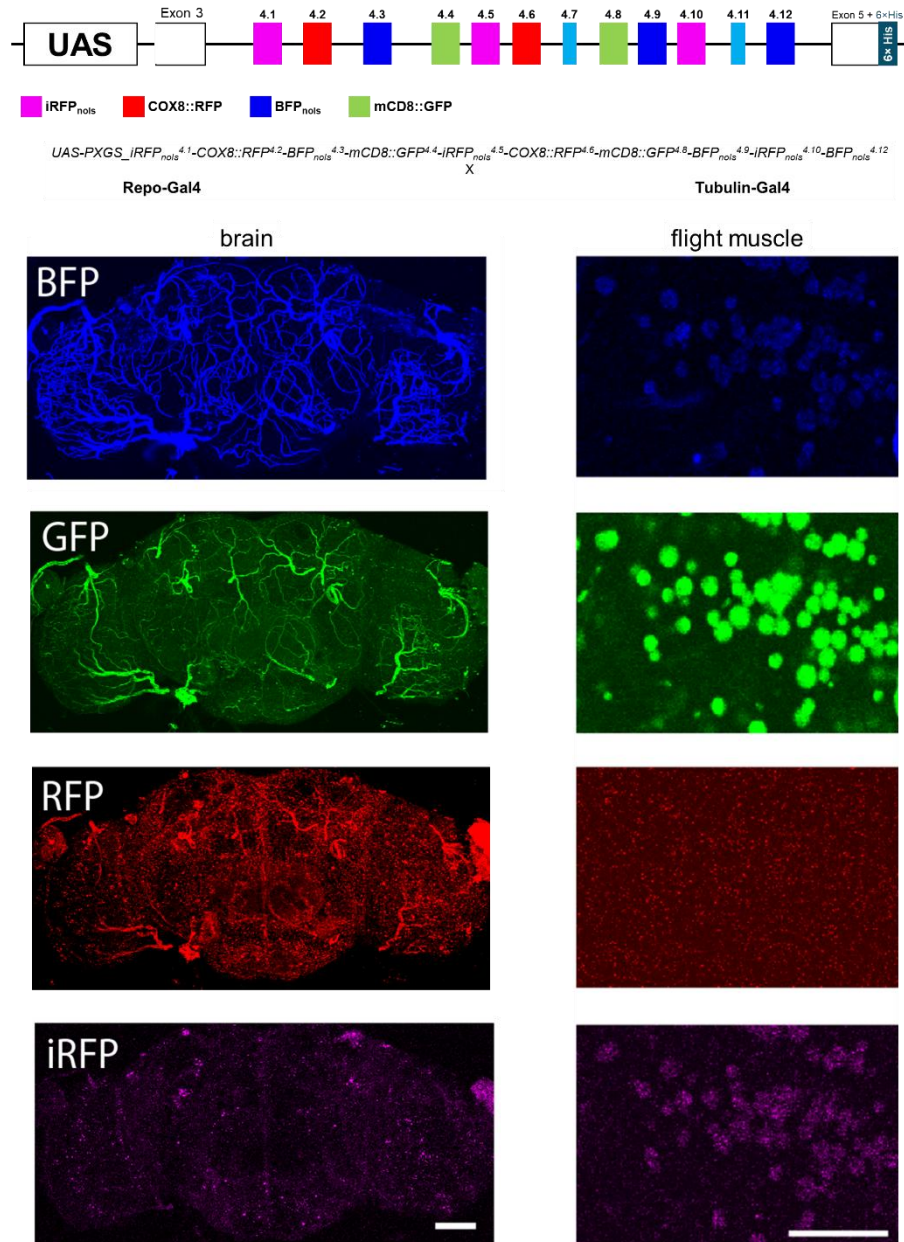


Figure 3, PXGS can drive expression in non-neural cells. Crossing the transgenic fly *UAS-PXGS_iRFP_{nols}^{4.1}-COX8::RFP^{4.2}-BFP_{nols}^{4.3}-mCD8::GFP^{4.4}-iRFP_{nols}^{4.5}-COX8::RFP^{4.6}-mCD8::GFP^{4.8}-BFP_{nols}^{4.9}-iRFP_{nols}^{4.10}-BFP_{nols}^{4.12}* to the glial *Repo-Gal4* driver (left column) or to the ubiquitous *Tubulin-Gal4* driver (right column) showed expression in non-neural cells. The subcellular localization of the fluorophores was observed in the flight muscle (right column), but was less apparent in glial cells in the brain (left column). Scale bars are 50µm.

Dotted lines mark the midline. Scale bars are 50µm. Statistical significance comparisons to control are indicated directly above the bar representing each genotype. No asterisks indicate the comparison was not significant. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$.

References

- Agmon, N., Pur, S., Liefshitz, B., and Kupiec, M. (2009). Analysis of repair mechanism choice during homologous recombination. *Nucleic Acids Res* 37, 5081-5092.
- Aigner, H., Wilson, R.H., Bracher, A., Calisse, L., Bhat, J.Y., Hartl, F.U., and Hayer-Hartl, M. (2017). Plant RuBisCo assembly in *E. coli* with five chloroplast chaperones including BSD2. *Science* 358, 1272-1278.
- Ban, N., Nissen, P., Hansen, J., Moore, P.B., and Steitz, T.A. (2000). The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science* 289, 905-920.
- Bhattacharya, A. (2009). Protein structures: Structures of desire. *Nature* 459, 24-27.
- Bieniossek, C., Imasaki, T., Takagi, Y., and Berger, I. (2012). MultiBac: expanding the research toolbox for multiprotein complexes. *Trends Biochem Sci* 37, 49-57.
- Brand, A.H., and Perrimon, N. (1993). Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* 118, 401-415.
- Brites, D., Brena, C., Ebert, D., and Du Pasquier, L. (2013). More than one way to produce protein diversity: duplication and limited alternative splicing of an adhesion molecule gene in basal arthropods. *Evolution* 67, 2999-3011.
- Celotto, A.M., and Graveley, B.R. (2001). Alternative splicing of the *Drosophila* Dscam pre-mRNA is both temporally and spatially regulated. *Genetics* 159, 599-608.
- Ceolin, S., Hanf, M., Bozek, M., Storti, A.E., Gompel, N., Unnerstall, U., Jung, C., and Gaul, U. (2020). A sensitive mNeonGreen reporter system to measure transcriptional dynamics in *Drosophila* development. *Commun Biol* 3, 663.
- Chen, B.E., Kondo, M., Garnier, A., Watson, F.L., Puettmann-Holgado, R., Lamar, D.R., and Schmucker, D. (2006). The molecular diversity of Dscam is functionally required for neuronal wiring specificity in *Drosophila*. *Cell* 125, 607-620.
- Coker, J.A., Katis, V.L., Fairhead, M., Schwenzer, A., Clemmensen, S.B., Frandsen, B.U., de Jongh, W.A., Gileadi, O., Burgess-Brown, N.A., Marsden, B.D., *et al.* (2022). FAS2FURIOUS: Moderate-Throughput Secreted Expression of Difficult Recombinant Proteins in *Drosophila* S2 Cells. *Front Bioeng Biotechnol* 10, 871933.

- Corrales, M., Cocanougher, B.T., Kohn, A.B., Wittenbach, J.D., Long, X.S., Lemire, A., Cardona, A., Singer, R.H., Moroz, L.L., and Zlatic, M. (2022). A single-cell transcriptomic atlas of complete insect nervous systems across multiple life stages. *Neural Dev* 17, 8.
- Cvetkovska, V., Hibbert, A.D., Emran, F., and Chen, B.E. (2013). Overexpression of Down syndrome cell adhesion molecule impairs precise synaptic targeting. *Nat Neurosci* 16, 677-682.
- Dong, H., Yang, X., Wu, L., Zhang, S., Zhang, J., Guo, P., Du, Y., Pan, C., Fu, Y., Li, L., *et al.* (2023). A systematic CRISPR screen reveals redundant and specific roles for Dscam1 isoform diversity in neuronal wiring. *PLoS Biol* 21, e3002197.
- Donnelly, M.L., Gani, D., Flint, M., Monaghan, S., and Ryan, M.D. (1997). The cleavage activities of aphthovirus and cardiovirus 2A proteins. *The Journal of general virology* 78 (Pt 1), 13-21.
- Doronina, V.A., de Felipe, P., Wu, C., Sharma, P., Sachs, M.S., Ryan, M.D., and Brown, J.D. (2008). Dissection of a co-translational nascent chain separation event. *Biochemical Society transactions* 36, 712-716.
- Dos Santos, J.V., Yu, R.Y., Terceros, A., and Chen, B.E. (2019). FGF receptors are required for proper axonal branch targeting in *Drosophila*. *Mol Brain* 12, 84.
- Gibson, D.G., Young, L., Chuang, R.Y., Venter, J.C., Hutchison, C.A., 3rd, and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* 6, 343-345.
- Gleizer, S., Ben-Nissan, R., Bar-On, Y.M., Antonovsky, N., Noor, E., Zohar, Y., Jona, G., Krieger, E., Shamshoum, M., Bar-Even, A., *et al.* (2019). Conversion of *Escherichia coli* to Generate All Biomass Carbon from CO₂. *Cell* 179, 1255-1263 e1212.
- Hausmann, I.U., Ustaoglu, P., Brauer, U., Hemani, Y., Dix, T.C., and Soller, M. (2019). Plasmid-based gap-repair recombineered transgenes reveal a central role for introns in mutually exclusive alternative splicing in Down Syndrome Cell Adhesion Molecule exon 4. *Nucleic Acids Res* 47, 1389-1403.
- Hong, W., Zhang, J., Dong, H., Shi, Y., Ma, H., Zhou, F., Xu, B., Fu, Y., Zhang, S., Hou, S., *et al.* (2021). Intron-targeted mutagenesis reveals roles for Dscam1 RNA pairing architecture-driven splicing bias in neuronal wiring. *Cell Rep* 36, 109373.
- Kays, I., Cvetkovska, V., and Chen, B.E. (2014). Structural and functional analysis of single neurons to correlate synaptic connectivity with grooming behavior. *Nat Protoc* 9, 1-10.

- Kozak, M. (1999). Initiation of translation in prokaryotes and eukaryotes. *Gene* 234, 187-208.
- Kreahling, J.M., and Graveley, B.R. (2005). The iStem, a long-range RNA secondary structure element required for efficient exon inclusion in the *Drosophila* Dscam pre-mRNA. *Mol Cell Biol* 25, 10251-10260.
- Kriz, A., Schmid, K., Baumgartner, N., Ziegler, U., Berger, I., Ballmer-Hofer, K., and Berger, P. (2010). A plasmid-based multigene expression system for mammalian cells. *Nat Commun* 1, 120.
- Lee, C., Kim, N., Roy, M., and Graveley, B.R. (2010). Massive expansions of Dscam splicing diversity via staggered homologous recombination during arthropod evolution. *RNA* 16, 91-105.
- Li, H., Janssens, J., De Waegeneer, M., Kolluru, S.S., Davie, K., Gardeux, V., Saelens, W., David, F.P.A., Brbic, M., Spanier, K., *et al.* (2022). Fly Cell Atlas: A single-nucleus transcriptomic atlas of the adult fruit fly. *Science* 375, eabk2432.
- Liu, Z., Chen, O., Wall, J.B.J., Zheng, M., Zhou, Y., Wang, L., Vaseghi, H.R., Qian, L., and Liu, J. (2017). Systematic comparison of 2A peptides for cloning multi-genes in a polycistronic vector. *Sci Rep* 7, 2193.
- Lo, C.A., and Chen, B.E. (2019). Parental allele-specific protein expression in single cells *In vivo*. *Dev Biol*.
- Lo, C.A., Kays, I., Emran, F., Lin, T.J., Cvetkovska, V., and Chen, B.E. (2015). Quantification of Protein Levels in Single Living Cells. *Cell Rep* 13, 2634-2644.
- Martin D. Ryan GL, L.E.H., Vanessa M. Cowton, Edwin Ten Dam, Xuejun Li, Michelle L. L. Donnelly, Amit Mehrotra, and David Gani (2002). The Aphtho- and Cardiovirus "Primary" 2A/2B Polyprotein "Cleavage". In *Molecular biology of picornaviruses*, E.W. Bert L. Semler, ed. (Washington, DC: ASM Press), pp. 213-223.
- Mizuguchi, H., Xu, Z., Ishii-Watabe, A., Uchida, E., and Hayakawa, T. (2000). IRES-dependent second gene expression is significantly lower than cap-dependent first gene expression in a bicistronic vector. *Mol Ther* 1, 376-382.
- Neufeld, S.Q., Hibbert, A.D., and Chen, B.E. (2011). Opposing roles of PlexinA and PlexinB in axonal branch and varicosity formation. *Mol Brain* 4, 15.
- Pelletier, J., and Sonenberg, N. (1988). Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA. *Nature* 334, 320-325.

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

Qiao, H.H., Wang, F., Xu, R.G., Sun, J., Zhu, R., Mao, D., Ren, X., Wang, X., Jia, Y., Peng, P., *et al.* (2018). An efficient and multiple target transgenic RNAi technique with low toxicity in *Drosophila*. *Nat Commun* 9, 4160.

Ralston, E., and Hall, Z.W. (1989). Intracellular and surface distribution of a membrane protein (CD8) derived from a single nucleus in multinucleated myotubes. *J Cell Biol* 109, 2345-2352.

Ryan, M.D., King, A.M., and Thomas, G.P. (1991). Cleavage of foot-and-mouth disease virus polyprotein is mediated by residues located within a 19 amino acid sequence. *The Journal of general virology* 72 (Pt 11), 2727-2732.

Sari, D., Gupta, K., Thimiri Govinda Raj, D.B., Aubert, A., Drncova, P., Garzoni, F., Fitzgerald, D., and Berger, I. (2016). The MultiBac Baculovirus/Insect Cell Expression Vector System for Producing Complex Protein Biologics. *Adv Exp Med Biol* 896, 199-215.

Schmucker, D., and Chen, B. (2009). Dscam and DSCAM: complex genes in simple animals, complex animals yet simple genes. *Genes Dev* 23, 147-156.

Schmucker, D., Clemens, J.C., Shu, H., Worby, C.A., Xiao, J., Muda, M., Dixon, J.E., and Zipursky, S.L. (2000). *Drosophila* Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. *Cell* 101, 671-684.

Shcherbakova, D.M., Baloban, M., Emelyanov, A.V., Brenowitz, M., Guo, P., and Verkhusha, V.V. (2016). Bright monomeric near-infrared fluorescent proteins as tags and biosensors for multiscale imaging. *Nat Commun* 7, 12405.

Stahl, K., Warneke, R., Demann, L., Bremenkamp, R., Hormes, B., Brock, O., Stulke, J., and Rappsilber, J. (2024). Modelling protein complexes with crosslinking mass spectrometry and deep learning. *Nat Commun* 15, 7866.

Suciu, R.M., Aydin, E., and Chen, B.E. (2015). GeneDig: a web application for accessing genomic and bioinformatics knowledge. *BMC Bioinformatics* 16, 67.

Wiederstein, M., Gruber, M., Frank, K., Melo, F., and Sippl, M.J. (2014). Structure-based characterization of multiprotein complexes. *Structure* 22, 1063-1070.

Xu, B., Shi, Y., Wu, Y., Meng, Y., and Jin, Y. (2019). Role of RNA secondary structures in regulating Dscam alternative splicing. *Biochim Biophys Acta Gene Regul Mech* 1862, 194381.

Yue, Y., Yang, Y., Dai, L., Cao, G., Chen, R., Hong, W., Liu, B., Shi, Y., Meng, Y., Shi, F., *et al.* (2016). Long-range RNA pairings contribute to mutually exclusive splicing. *RNA* 22, 96-110.

Supplemental Information

PXGS: a Poly-Transgene Expression System based on Mutually Exclusive Splicing of *Dscam*

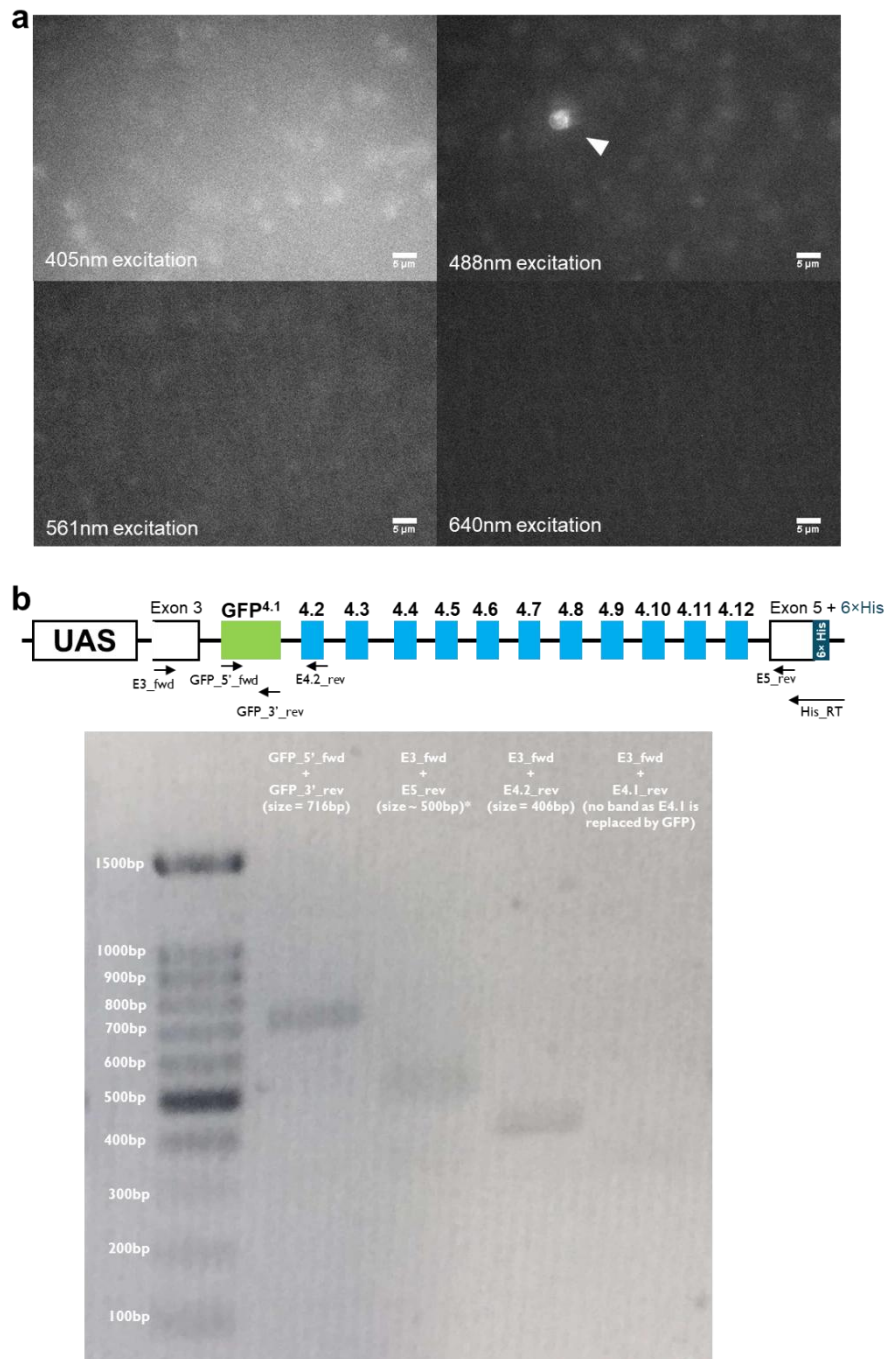
Renee Yin Yu^{1,†}, Alyeri Bucio-Méndez^{1,†}, Brian E. Chen^{1,2,*}

¹ Centre for Research in Neuroscience, Research Institute of the McGill University Health Centre, Montréal, Québec, Canada, H3G 1A4.

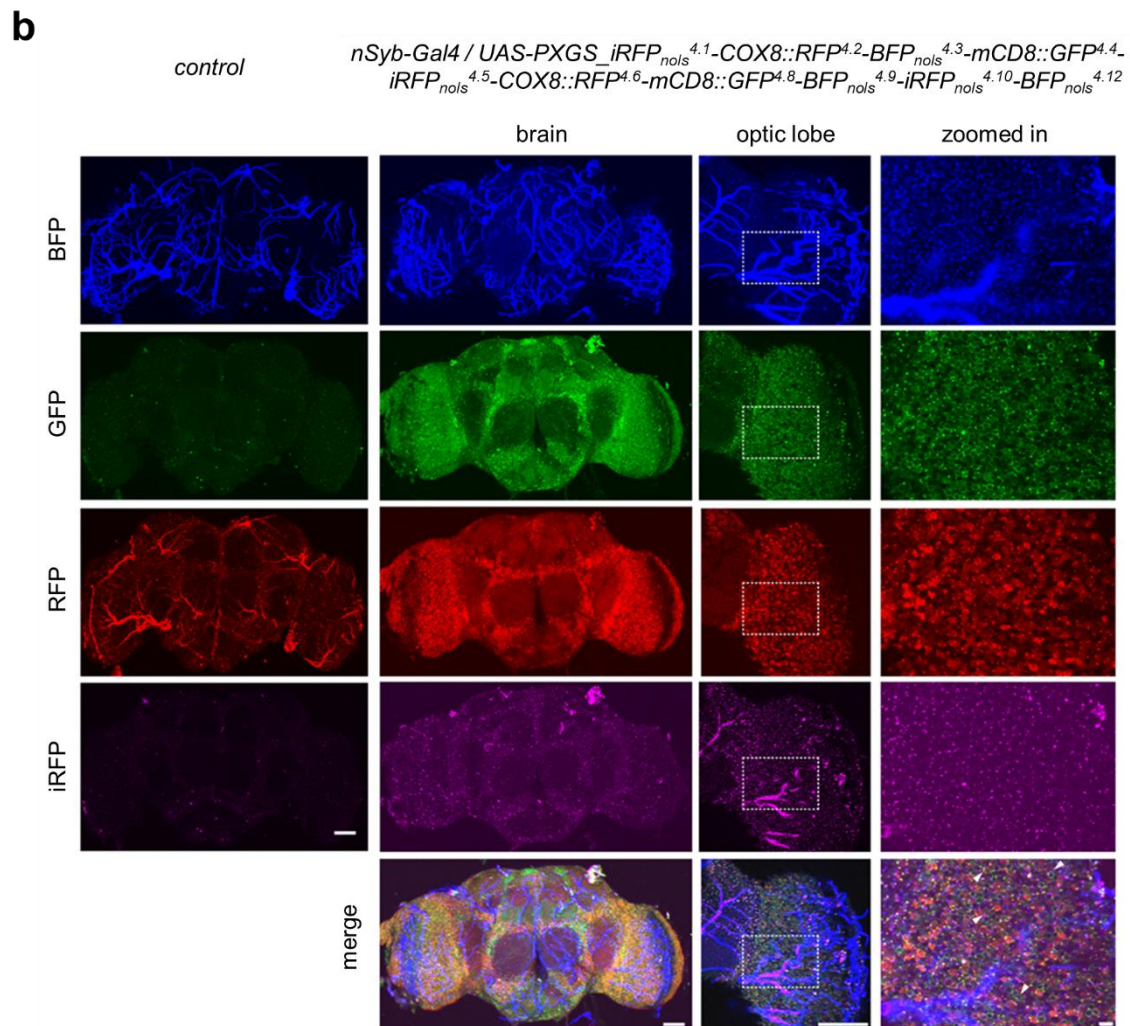
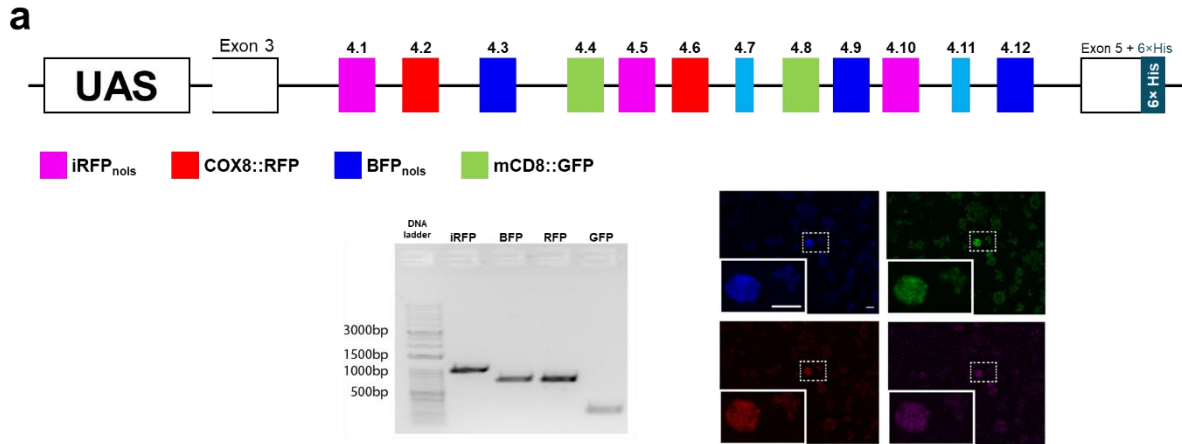
² Departments of Medicine and Neurology & Neurosurgery, McGill University, Montréal, Québec, Canada, H3G 1A4.

[†] These authors contributed equally to this work.

* Corresponding author: Brian E. Chen, Montreal General Hospital, 1650 Cedar Ave, L7-224, Montréal, Québec, Canada, H3G 1A4, tel: 514-934-1934 x42379, fax: 514-934-8216, brian.chen@mcgill.ca

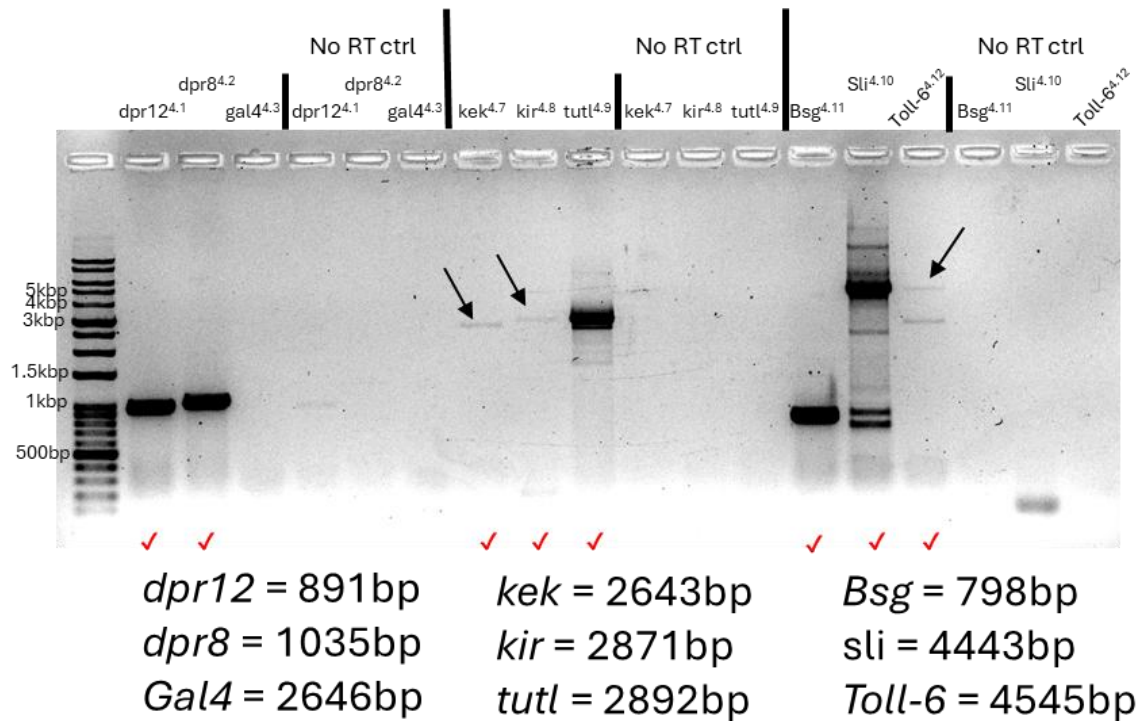


Supplemental Figure 1 – Complete replacement of an Exon 4 alternate can properly splice and produce functional protein. **a**, Replacing Exon 4.1 with the gene for *GFP* produces green fluorescence when co-transfected with *Actin5C-Gal4* in *S2* cells. Blue (upper left), green (upper right), red (lower left), and far red (lower right) channels are shown. Scale bars are 5μm. **b**, PXGS-specific RT-PCR of different exon 4 variants in transfected *S2* cells verified their appropriate splicing and expression. *GFP* mRNA band is expected at 716 bp, the remaining possible exon 4 variants are expected at ~500 bp (denoted with a *), Exon 4.2 is expected at 406 bp, and Exon 4.1 is not expected.

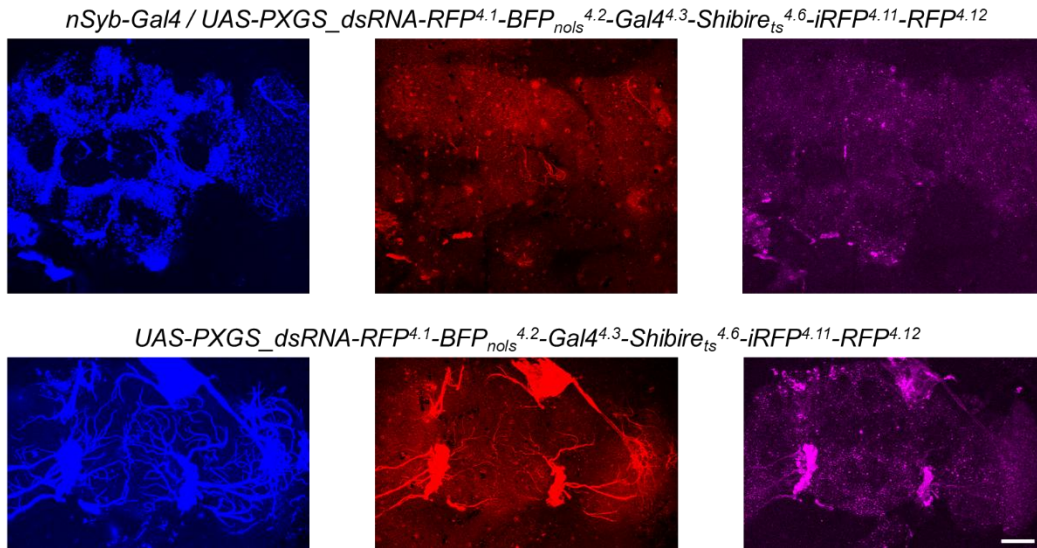
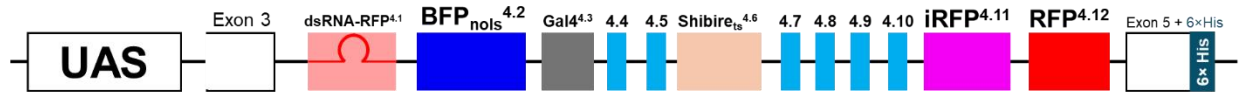


Supplemental Figure 2 – Replacing 10 of the 12 Exon 4 alternates with fluorescent protein transgenes increases the expression pattern in neurons. **a**, Co-transfection of *UAS-PXGS_iRFP_{nols}^{4.1}-COX8::RFP^{4.2}-BFP_{nols}^{4.3}-mCD8::GFP^{4.4}-iRFP_{nols}^{4.5}-COX8::RFP^{4.6}-mCD8::GFP^{4.8}-BFP_{nols}^{4.9}-iRFP_{nols}^{4.10}-BFP_{nols}^{4.12}* with *Actin5C-Gal4* in *S2* cells produced mRNA

at the expected sizes (*iRFP* = 1044bp, *BFP* = 810bp, *RFP* = 804bp, *GFP* = 270bp), and produced fluorescent cells. Scale bars are 10µm. **b**, Crossing the pan-neuronal *nSyb-Gal4* driver line with the transgenic fly *UAS-PXGS_iRFP_{nols}^{4.1}-COX8::RFP^{4.2}-BFP_{nols}^{4.3}-mCD8::GFP^{4.4}-iRFP_{nols}^{4.5}-COX8::RFP^{4.6}-mCD8::GFP^{4.8}-BFP_{nols}^{4.9}-iRFP_{nols}^{4.10}-BFP_{nols}^{4.12}* resulted in all neurons expressing all four fluorophores at higher levels than the four fluorophore PXGS fly. The differential subcellular localizations of the fluorophores can be observed at higher magnifications in the optic lobe. White arrowheads in the zoomed in merged image (bottom right) point to cells that express all four fluorophores evenly. Scale bars are 50µm, 50µm, and 5µm respectively from left to right.

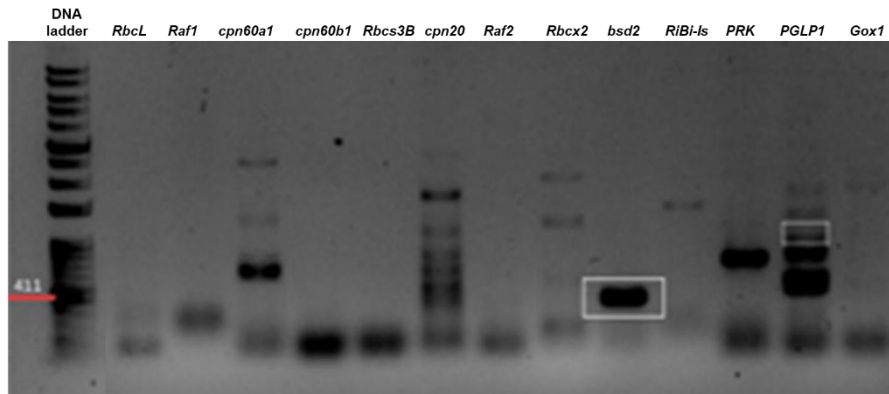


Supplemental Figure 3 – PXGS can express multiple cell surface receptors in *S2* cells. *UAS-PXGS_dpr8^{4.1}-dpr12^{4.2}-Gal4^{4.3}*, *UAS-PXGS_kek1^{4.7}-kirre^{4.8}-tutl^{4.9}*, and *UAS-PXGS_Toll-6^{4.10}-Bsg^{4.11}-sli^{4.12}* were each co-transfected into *S2* cells with *Actin5C-Gal4*. RT-PCR on each gene was performed 48 hours after transfection. The PCR products for *UAS-PXGS_dpr8^{4.1}-dpr12^{4.2}-Gal4^{4.3}* are shown in the left side, *UAS-PXGS_kek1^{4.7}-kirre^{4.8}-tutl^{4.9}* in the middle, and *UAS-PXGS_Toll-6^{4.10}-Bsg^{4.11}-sli^{4.12}* on the right. The expected sizes for each gene's PCR product is shown. *Gal4* mRNA from PXGS expression was not detected likely due to a mutation within the upstream intronic sequence that may interfere with its splicing.



Supplemental Figure 4 – PXGS can express dsRNA for RNAi knockdown. Flies expressing dsRNA against RFP did not have red fluorescence, but still expressed BFP. The low iRFP signal is possible due to the off-target effects of the RNAi. Scale bar is 50 μ m.

Actin-Gal4 / UAS-PXGS_ rbcL1-PQR-raf1^{4.1}-cpn60a1^{4.2}-cpn60b1^{4.3}-rbCS3B^{4.4}-cpn20^{4.5}-raf2^{4.6}-rbcx2^{4.7}-bsd2^{4.8}-RiBi-Is^{4.9}-PRK^{4.10}-PGLP1^{4.11}-GOX1^{4.12}



Supplemental Figure 5 – PXGS can express 13 large genes to re-create synthetic biology pathways. Actin-Gal4 flies were crossed to PXGS flies (*UAS-PXGS_ rbcL1-PQR-raf1^{4.1}-cpn60a1^{4.2}-cpn60b1^{4.3}-rbCS3B^{4.4}-cpn20^{4.5}-raf2^{4.6}-rbcx2^{4.7}-bsd2^{4.8}-RiBi-Is^{4.9}-PRK^{4.10}-PGLP1^{4.11}-GOX1^{4.12}*) expressing 13 genes to synthesize the carbon fixing enzyme, Ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) (Aigner et al., 2017; Gleizer et al., 2019). RuBisCO fixes carbon from atmospheric carbon dioxide to convert it into fuel, such as glucose. The two genes, *rbcL1* and *raf1* were inserted into position 4.1 with a CHYSEL *Protein Quantitation Reporter (PQR)* linker specific to *Drosophila* (Lo et al., 2015) demonstrating that a poly-cistronic RNA (the *PQR* linker) within the poly-cistronic PXGS can generate even larger numbers of transgenes. RT-PCR verified that all 13 genes were expressed in the flies. Thus, PXGS can be used to express complex enzymatic or molecular pathways in *Drosophila in vivo*.

PXGS_empty_annotated	TGCCGACCAAAAAGGACCCGTCTTCCTCAAGGAACCCACCAACCGCATTGACTTCTCCAACCTCCACGGGCGCAGAGATCGAGTGCAAGGCCAGCGGCAATCCCTtTGCCCCGAGATTATTTGGATCAGGAGCGACGGTACCGCCGTGGGTGATGTGCCCGGATTGCGTCAGATCTCATCCGACGGCAAGCTGGTCTTCCCTCCATTCCGCGCCGAGGACTACCGCCAGGAGGTCCATGCCCAGGTGTACGCCTGCCTaGCCCCGCAACCAGTTCGGATCCATTATCTCCCGGGACGTCCtTGTCGAGCCGTAAGTTGTCAGGGTTCAGAGGTAAACCAGGGGCAGTGCAAA
Exon3	GTAATGTTGGTAATTGACAAAGAAAGTCATTCTTAAAAAAGGATAGATATGAGCTTTATATATTTTTAAAAAGTTTAAAAAAATATTTAATAATAATATATAAAATATTACAAATATTCATTTAAATGAAACTCAACATTTAGTTTAAAGATTTTAAATACAATCTGCAATCAAAATACAAGAATAGAAGATGAA
Exon 3 mutations:	AAAGCTGCGAGTTAAAGTAACTAATTATTTGAAGTCGGTAGAAAATTCCTTTGAATTAGGCAACCAACAGGTGACTACTTTGCACTGCCCTGAGTTTACCTTTCTTTACAGGTTACCCCAAATGAAGGCCTATGTATCTGCGTTGGTTTCGTATCTCCACTACTTGGCACTTTCAGTTTGCAGCTAAAGTTGCCGGTTGCCTTTGCGTGTCAAGTCGCTCTCTGCAGTCTTTAACCCCTTGTTTGACTTATTGTTGCACCGGCTGCGGCAATTAAACCGAGTTTTACGGGATTTCATCGAACGCTCTGCGAATTGCCAAAGTTTTCCGCACGCTTTCATTTTACACATTTCCTAGTCGAACTATGCAACGTTTAGAGGAGTTTAATGCAAGTTGGCATTTCAGTTTAATCGGTTGTTTTCGCAGCTTTACTTCCTGTGCGCAATGTTTTGTGCGGCAACTGTACCTGTCATGCCCGTGCATGTCTGTTTCAGCAACCTTTAACAGCTTCCCTGGAAGATGTGGCAGCTGCGACGATGACGATGATGAAGAAGATGATGAAACTGTCAGACTATTTAGCGCACACCGTGCGGTTGACGGGTTGCCAGCTGATCGAGTTATGGACCAGACCCAACTCCCTTGGAATCAGTCACTTCCGCTTCCATGGAGCCATCCCTAAATGGGCCAGACTTGAGCACATTTTAGCTAATTATGTCGGAGCCATGCCACCCGCTTACTCATCGCCAGCTGACGTGTTCCCTGGCAATTAGCCAGCACACCTATCCCAGACACCACCTCTCGACCCCGTTCCAGAAGCCTGACATCATTATTGTCGTTTGCTTTGCTTATTAACCGGAGGAATTTATCATTTTATAGTATGCTTATGTGTGAATAAGGCGCAGACGAGGAAAACATTCTGCACAGCCGGGCCGCGAGCTAGTTGGCCGAAATGAGTTATGTCTACACCCTGAACACGGGCTGGCGTAAAGAATATTTTGGTTAATTGAAATTAACTTTTGCCTGTCAAGCAGCGGCGCGATGATAGTTTCGGGGTTTTCGACCAGAAGGTGCTTTACAACGGCAGAATTTTCCAGGAATCAGTTCCTAAGCACTTTCAAGGTTGTGAGAATGTTTTGGACATAATAGGTAATAATCGGCCTTTTCCCAGTGGTTGCCAGTACTACGAGGCCGATGTTAAACAAGGAGCACGTTATAAGAGGCAATTCGGCGGGTCA
A -> t: get rid of start	
T -> a: create a stop	
A -> t: get rid of start	
Exon 4.1 to Exon 4.12	
Exon 5	
His tag	

TCAAGTGTCTGATTCCATCCTTCGTGGCCGATTTCGTCTGAAGTGGT
GTCCTGGCACACCGATGAGGAGGAGAACTACTTTCCGGGCGCG
GAATACG GTGCGATCCGAGCTGCAGATGCAGGCAGATAGGATAC
ATGTGTAGTTTTGCTCTACTAAAAGCATGCTGGGGTCGCGATTCCG
CTAGTTGCACAGAAAATCCCTGCATTCCATCATTCAATTCCTTAT
TTAATCTGAATCTCACCCTATTATCGTTTCTCCTACCTGTTTAGTCG
TTTCCCAACACTACGAAGAAGATATACACAAGGCATTTGTCATCCG
CGGCAATTCCGCCATATTGAAATGCGATATACCGTCGTTTGTAGCC
GATTTTCGTGAATGTTATATCCTGGCATAGTGACGAGAAAGAGAATTT
CTATCCAGGCACTGAATATG GTGTGTGACTAGAAGAACAGTCTCTGT
TTGTGAGCTTTTAGAAAATATCCTTAAACTACCTTAACCGCTACTGT
ACAACACTGCAAAAAAGGCAAACTCACTGAAACACTTGTGCACAG
TGCAAGGGCAGTTAACGGGAAACACTAATTACAAAATTCTGTAAAA
ATACAACACACCAAACCTCAAAGAATAACGTATTAACCAACACTGGC
TAGGCAGAAACCAGTCGAGAATCCTTCGCTCCCTCGTCGTAGCTT
TCCAATTCGTCAGAGTTTCGTTTTAGTTCTACCCAGAATTTGATT
CCTTCGGTTTTATTACATAAACGCATTTATTGATTGCTGTCATTGCT
GTTTTAGTCGTTGCCCAATACTACGACACCGATGTAAATAAGGCCTA
TGTAATACGCGGCAATGCGGCGTTCTGAAATGCGAAATCCCCTC
CTTTGTGGCCGATTTTGTGAGCGTTGTCTCCTGGCACACCGATCAG
AACGAGGACTTTCTGCCAGGCAGCGAATACG GTTTGTTGGCCAAC
TTTTGATTACATTTTGGGCTAGCTATACATACATAGGGTTTGTACAGT
TGCGTGCCACCTAAGAATTATGGATAAGTACCTAACCATTTTCCCTC
CTGCATCTATTATGTTTACTTTTATGCGTTTATCTTCGCTATCTCGGGT
GACTGCGTGTTGATGAGCCATGTTATCAAAACCATTATTAACCTCTT
GCCCCGAGATTTTCTCAGGGAAGCGCGGAGCTCGCGAGCTGATT
GCGCTTAGGGTCCCATTAAGTGCACTCCGCGATATTGTCTGCGCT
GTTAGAAGACACTCTCGTTTCAATTAGCTAGCTTACATTTTACAC
GCCCCGCCGCACGCAAGGCTTGGACTTAACTTGGGTTCTGTAGG
CGCCCCATCCCCCATCTTCCCTCCGTACCATCGCATCATTATGTA
ATCAGTGGCTATGTGAATCACCTTCAG TGTGCAGCAGTTCTATGA
ATCGGAGGTCAACAACGAGTACGTATAAGGGGCAATGCAGCGG
TGCTGAAGTGCTCGATTCCATCCTTTGTTGCGGACTTTGTGCAGGT
CGTATCCTGGCAGGACGAAGAGGGTCAGCTCTACGGCTCGCTAG
GCGATCAACAGGGAACGG GTACACATGCACTATAGCGAAACCTAA
TGGTTATGTAGGATTATCCACAGCCCATCCCGCACCTACCCCAAT
CCTTCATTAGCAATTTCAATTCATTTTGGTTCCTACTCATAAAAACTC
TTGCTTTACAGTCGTGAATCAGTATTACGAGGCTGAGGTTGTTCCG
AGTACGTTATCCGTGGCAATACGGCTGTATTGAAGTGCACTATTCC
CAGTTTTGTAGCCGACTTCCTGCGGGTCGAGGCCTGGGTGGCCA
GTGACGGCACTGAGTACGCTCCGGAAGAGGATTTG GTACAGTCC

CGCACCGTTGAAACTATTAAGAAGTTAGTCCTAAGCAGCAATCCCA
TTCCTTCCCCATTTCAATTTGCTTCAGCTTCTGTTTCATTTCAACCACT
CTACCCCATTTTAAATCGCAGTTGTGAATCAGTTTTACGGCGCCGAT
ATCCTGATGGAGTATGTCATCAGGGGAAATGCGGCGGTTTTGAAAT
GTTCCATTCCGTCGTTTCGTGGCCGATTTTGTGCGTGTGAGTCCTG
GATAGATGAGGAAGGCACCGAACTGCGTCCCTCTGAGAACTATGG
TATTACCGCAGAGTCAAAAATTCTTGCAACGGCAACACTTTGTGGG
GGTTTCGTTCCCTATCCCATCCTATCTCCAACCTTTTCTATATCTTCCGT
TATTGCACACCTTTGCAGTTGTCAATCAGTTCTATGAGGCGGAGATT
ATGACCGAGTATGTTATCAGGGGTAATGCTGCCGTTCTGAAGTGTC
CATCCCGTCGTTTCGTGGCTGATTTTGTTCGAGTTGAGTCCTGGATC
GATGATGAGGGCAATGTCTTGTCCTTTTCGGACAATTACGGTGCGA
GGCAAAGATGTTTCATCCAGTAGTTTATGCTTAATTTCCACCAGCTC
CATTCCCAGTGATCCCCTCTTTTCCATTTACTTATTATTTATTTCGATT
CAAAGTCGTTTCACAGTTCTACATAACCGAGGCCGAGAATGAGTAC
GTAATTAAGGGAACGCGGCGGTGGTCAAGTGTAAGATCCCATCC
TTCGTGGCCGATTTTCGTGCAGGTTGAGGCCTGGGTGGACGAGGA
GGGCATGGAGTTGTGGCGCAACAATGCCACCGACGCCTATGGTA
CAGTTCAAGCAAAACATACTGCTTCGGTTAAGTGCACTAGGTC
CCAACAACCACTTCCTTGACTCCCAACCCCCAAACCCATATTTTA
CACTACATTATTCCTTTGTTTTTCTATCGACTCAGTGGTCATCCAAAG
CTATGAATCGGAGGCGGACAATGAGTACGTCATTAGGGGGCAACTC
TGTGGTGATGAAGTGCGAGATTCCCTCCTACGTGGCCGACTTCGT
GTTTCGTGATCTCTGGCTGGACTCGGAGGGTCGCAACTACTATCC
GAACAATGCCGCAGAGACGGGTACTTGCCAGGTTTCTTTGCCGT
AGCTCTCTCAGGCTCCTCTATAGCTATCCAGAAACCCATTCCCCTT
CACTTCGGTTATTATTTACCAAGTTCGCTCTGATTTCCCTCAGTTGT
CCACCAGTTCTACCAGACACGGGTCATCGATGAGTTTGTGCTGCG
TGGCAACTCCGCCACCTTGAAGTGCTTGGTGCCCTCGTTTGTGGC
AGACTTCATCGATGTGAGGGTTGGATCGACGAGGAGGGCGTGG
AGATCCTGCGGCCCTCCCGGACGACTCCGTTGGTAATCACTAC
GACTTTCGGCATTAGCATAGTTAGAGTACGAAACCTTCAACCCACC
GACCCTTCCCTGTCCCGATTTTTAGTTTCTACCTCCCGTCTTGCA
GTGGTGAAGCAGTTTTTCGAGTCGCAAGTCTACGACGAGTATGTGA
TCAAGGGCAATGCGGCCATTTTCAAGTGCCAGACGCCCTCCTTGG
TGGCGGATCACATAGACATAACCGACTGGATCGATACCGAGGGC
GAGGTCTTCACGAAGAACATAACTGTAGGAGTCCGAGTCTCGTA
GATTTGTGCCGTTTTCTCAGTTACATGAATGGGGATCCTCTTTCAGA
CGCCTCTTTTATCCTTATGATTTAACCCACGTCCCACTCACTCAAT
CAATGTCTTCGTACACTTTGCAGTTGTGCCACAATCATACACCGTCA
ATGTCATGGACGAGTCCATACTGCGAGGCAATAGCGCCATCCTAA

	AGTGCCACATTCCCAGTTTTGTGGCCGATTCATAGTGGTCGATTC GTGGGTGGAGGACGAGGAGCGAGTTATCTATCCCCAGGAGGATAT CGCGGAAAGCGGTAAATTTACAGACTAAGAGCTCTCGCATTGCCC ACTCCTCCGTAACATCATCCTGCGTCCCCTATCCAATTCGGTTCGC TTCGATCTGGTCTCGATGTGGTTTCTTTAGAGCTCTGTAAAAGAAGTT CTCCACGTTACGCGCTGCTTAGGCTTTTGAATGGCCTTCATTGTGT ATTTCTTTCATTCTGTCCGCCTTTTCTTTCTATGTTTTTGTGTTTTTTT TGTTTTCCCATCCCAACCCTTTTGTGTTTTCCCTCCAAAAATCCAT GGGTCGCTCCTGACTTGCCCCAAAAGAAGCCGAAACCCCTTCAT GTAGTTGAACGTAGCGTTGATAAAGTAGCGTAGTGTAATAAATGCA AAAATGCTAATAACCCATCGACCACGTAACAATATGTAAATACATATC ATTTCAAACGAATAATCCAACATTGCTGTAAATTGTATATACCATAAAGA TTCGAGTATATATAATATGTAAACATGCATTTTCATCTAAGATTATTGGCAT TTTTGTTTCATGCATGTCTAGAATTAAGCACAATGCCATGGAATTTGG AGAATCACGATCGAAGCAATCAACTAAAACCTATTTGAATTGTAAAAAT GTATCTCAATTCGAAATTTAAATAGATTACAATCAATCTTCCGGTTTG ATTCTGTTTGTTTATACTACAAACTCTTGGGCATTGTTGGTGTATAGTA AATGTTTATGGAACCTTTGTTAAAACAAATTCGATTGAAATCAAAATCAA AGACTAAAGGAGATCCTAACAAGCTTAGTAGTTAATCTTGCTCTAAA AAATTTCATAAATGAATCAAATTAATAACGAATTCTATTTTTAATCAAAC ATGTGCGAAATAACTACAAGGGCTTCTCATTTAGTCTATAGATCTTAA CTTATCCTTAATCATTTCAAAGTCACATTGCATGGTCAACGATTAGCA GCATTCAATTTTTGCACAATTAAGTAACACAAAATGAAAAATGATT ACCAGCCATGTGGCTTTCAAATGTAGAGCGAATTAATTCAATTGAAT GCATGCTAGTCAATTGATTTTGTTGCGCGTACTGTATTGTTAATAAGTT AAGTTAGCCGACAATTCATCACTAACATATTCTTTATTCTCTATCTGCA CATATCAAATATCAGATGGAAAGTACCTGGTATTGCCCTCTGGAGA GCTGCACATCCGTATCATCACCATCACCACATAA
--	--

Supplemental Table 1 – DNA sequence of PXGS construct with mutated Exon 3, endogenous Exon 4, and 6×Histidine tagged Exon 5.