Natural variation in stochastic photoreceptor specification and color preference in *Drosophila*

Caitlin Anderson[1], India Reiss[1], Cyrus Zhou[1,*], Annie Cho[1,*], Haziq Siddiqi[1,*], Ben Mormann[2], Cameron M. Avelis[3], Alan Bergland[4], Elijah Roberts[3], James Taylor[1], Daniel Vasiliauskas[5], and Robert J. Johnston Jr.[1#]

1. Department of Biology, Johns Hopkins University, 3400 N. Charles Street Baltimore, MD 21218-2685, USA

2. Center for Developmental Genetics, Department of Biology, New York University, 100 Washington Square East, New York, NY 10003-6688, USA

3. Department of Biophysics, Johns Hopkins University, 3400 N. Charles Street Baltimore, MD 21218-2685, USA

4. Department of Biology, University of Virginia, 063A Gilmer Hall, Charlottesville, VA 22903, USA

5. Centre National de la Recherche Scientifique-Unité Mixte de Recherche 7592, Institut Jacques Monod, Université Paris Diderot, Paris, France

Corresponding author. Email: robertjohnston@jhu.edu

* Authors contributed equally

Key words: stochastic cell fate specification, color preference, photoreceptor, natural variation, Spineless, Klumpfuss

Abstract

Each individual perceives the world in a unique way, but little is known about the genetic basis of variation in sensory perception. Here we investigated natural variation in the development and function of the color vision system of *Drosophila*. In the fly eye, the random mosaic of color-detecting R7 photoreceptor subtypes is determined by stochastic expression of the transcription factor Spineless (Ss). Individual R7s randomly choose between Ss^{ON} or Ss^{OFF} fates at a ratio of 65:35, resulting in unique patterns but consistent proportions of cell types across genetically identical retinas. In a genome wide association study, we identified a naturally occurring insertion in a regulatory DNA element in the *ss* gene that lowers the ratio of Ss^{ON} to Ss^{OFF} cells. This change in photoreceptor fates shifts the innate color preference of flies from green to blue. The genetic variant increases the binding affinity for Klumpfuss (Klu), a zinc finger transcriptional repressor that regulates *ss* expression. Klu is expressed at intermediate levels to determine the normal ratio of Ss^{ON} to Ss^{OFF} cells. Thus, binding site affinity and transcription factor levels are finely tuned to regulate stochastic on/off gene expression, setting the ratio of alternative cell fates and ultimately determining color preference.

Organisms require a diverse repertoire of sensory receptor neurons to perceive a range of stimuli in their environments. Differentiation of sensory neurons often requires stochastic mechanisms whereby individual neurons randomly choose between different fates. Stochastic fate specification diversifies sensory neuron subtypes in a wide array of species including worms, flies, mice, and humans (Ressler et al. 1993; Roorda and Williams 1999; Troemel et al. 1999; Hofer et al. 2005; Johnston and Desplan 2010; Magklara and Lomvardas 2013; Alqadah et al. 2016; Viets et al. 2016). How naturally occurring changes in the genome affect stochastic mechanisms to alter sensory system development and perception is poorly understood. To address this question, we investigated natural variation in stochastic color photoreceptor specification in the *Drosophila* retina.

The fly eye, like the human eye, contains a random mosaic of photoreceptors defined by expression of light-detecting Rhodopsin proteins (Montell et al. 1987; Bell et al. 2007; Johnston and Desplan 2010; Viets et al. 2016). In flies, the stochastic on/off expression of Spineless (Ss), a PAS-bHLH transcription factor, determines R7 photoreceptor subtypes. Ss expression in a random subset of R7s induces 'yellow' (**y**R7) fate and expression of Rhodopsin4 (Rh4), whereas the absence of Ss in the complementary subset of R7s allows for 'pale' (**p**R7) fate and Rhodopsin3 (Rh3) expression (**Fig. 1A**)(Wernet et al. 2006; Johnston et al. 2011; Thanawala et al. 2013; Johnston and Desplan 2014). The on/off state of Ss in a given R7 also indirectly determines the subtype fate of the neighboring R8 photoreceptor. **p**R7s lacking Ss signal to **p**R8s to activate

expression of blue-detecting Rhodopsin 5 (Rh5). **y**R7s expressing Ss do not send this signal, resulting in expression of green-detecting Rhodopsin 6 (Rh6) in **y**R8s (**Fig 1A**)(Franceschini et al. 1981; Montell et al. 1987; Zuker et al. 1987; Chou et al. 1996; Huber et al. 1997; Chou et al. 1999; Mikeladze-Dvali et al. 2005; Jukam and Desplan 2011; Hsiao et al. 2013; Johnston 2013; Jukam et al. 2013; Jukam et al. 2016).

The stochastic decision to express Ss is made cell autonomously at the level of the *ss* gene locus via a random repression mechanism. The *R7/R8 enhancer* induces *ss* expression in all R7s, whereas two silencer regions (*silencer 1* and *2*) repress expression in a random subset of R7s (**Fig. 1B**)(Johnston and Desplan 2014).

Though the stochastic expression of Ss is binary (i.e. on or off) in individual R7s, it does not result in a simple 50:50 on/off ratio across the population of R7s in a given retina. In most isogenic lab stock flies, Ss is on in ~65% of R7s and off in ~35% (**Fig. 1C**)(Wernet et al. 2006; Johnston and Desplan 2014). Here, we find that the proportion of Ss^{ON} to Ss^{OFF} R7s varies greatly among fly lines derived from the wild. We performed a genome-wide association study (GWAS) and identified a single base pair insertion that increases the affinity of a DNA binding site for a transcriptional repressor, significantly reducing the Ss^{ON}/Ss^{OFF} ratio. This genetic variant changes the proportion of photoreceptor subtypes and alters the innate color preference of flies.

sin decreases the ratio of Ss^{ON} to Ss^{OFF} R7s

To determine the mechanism controlling the ratio of stochastic on/off Ss expression, we analyzed the variation in 203 naturally-derived lines collected from Raleigh, North Carolina (Drosophila Genetic Reference Panel (DGRP))(Mackay et al. 2012). We evaluated Rh4 and Rh3 expression as they faithfully report Ss expression in R7s (i.e. Ss^{ON} = Rh4; Ss^{OFF} = Rh3) (**Fig. 1A**)(Thanawala et al. 2013; Johnston and Desplan 2014). To facilitate scoring, we generated a semi-automated counting system to determine the Rh4:Rh3 ratio for each genotype (**Fig. 1C**).

To assess the variation in the DGRP lines attributable to the *ss* locus and limit the phenotypic contribution of recessive variants at other loci, we crossed each DGRP line to a line containing a ~200 kb deficiency covering the *ss* locus and analyzed Rh3 and Rh4 expression in the F1 male progeny (**Fig. 1D**). This genetic strategy generated flies hemizygous (i.e. single copy) for the wild-derived *ss* gene locus, heterozygous wild-derived/lab stock for the 2nd, 3rd, and 4th chromosomes, and hemizygous lab stock for the X chromosome (**Fig. 1D**). While the lab stock expressed Ss (Rh4) in 62% of R7s under these conditions, expression among the DGRP lines varied significantly, ranging from 19 to 83% Ss^{ON} (Rh4) (**Fig. 1E-F; Supplemental Table 1**).

To identify the genetic basis of this variation, we performed a genome-wide association study (GWAS) using the Ss^{ON} (Rh4) phenotype data and inferred full genome sequences of the progeny of each DGRP line crossed with the *ss* deficiency line. We performed an association analysis and identified a

5

single base pair insertion within the *ss* locus ("*ss* insertion" or "*sin*") that was significant ($p < 10^{-13}$) after Bonferroni correction (**Fig. 1G**). *sin* was enriched in DGRP lines with a low ratio of Ss^{ON} to Ss^{OFF} R7s (**Fig. 1F and H**).

We next confirmed the regulatory role of *sin*. Naturally derived lines from Africa that are homozygous for *sin* displayed a decrease in the proportion of Ss^{ON} (Rh4) R7s compared to lines from Africa lacking *sin* (**Fig. 1I**)(Lack et al. 2015). We identified *sin* on a balancer chromosome (*TM6b*) in a lab stock that similarly displayed a decrease in the proportion of Ss^{ON} (Rh4) R7s when *ss* was hemizygous (**Fig. 1J**). To definitively test the role of *sin*, we used CRISPR to insert *sin* into a lab stock. Flies hemizygous for CRISPR *sin* alleles displayed a significant decrease in the proportion of Ss^{ON} (Rh4) R7s (**Fig. 1K**). Thus, *sin* causes a decrease in the ratio of Ss^{ON} to Ss^{OFF} R7s.

sin shifts innate color preference from green to blue

As *sin* alters the proportion of color-detecting photoreceptors, we hypothesized that it would also change color detection and preference. When presented with two light stimuli in a T-maze (Tully and Quinn 1985), flies will phototax toward the light source that they perceive as more intense (**Fig. 2A**)(McEwen 1918; Heisenberg and Wolf 1984; Choe and Clandinin 2005). The absorption spectra of Rh3 and Rh4 significantly overlap in the UV range (Feiler et al. 1992), complicating behavioral assessment of color preference caused by differences in R7 photoreceptor ratios. Instead, we focused on the perception of blue light by Rh5 and green light by Rh6 in the R8 photoreceptors, as these

6

Rhodopsins have more distinct absorption spectra (Salcedo et al. 1999). Because R8 fate is coupled to R7 fate(Chou et al. 1996) (**Fig. 1A**), we predicted that flies with *sin* would have a low ratio of Rh6- to Rh5-expressing R8s and would consequently prefer blue light, while flies without *sin* would have a higher ratio of Rh6- to Rh5-expressing R8s and would instead prefer green light. Indeed, DGRP lines containing *sin* preferred blue light, while DGRP lines lacking *sin* preferred green light (**Fig 2A-C; Supplemental Table 2**), showing that *sin* changes innate color preference in flies.

sin increases the binding affinity for the Klumpfuss transcription factor

sin is a single base pair insertion within a previously uncharacterized non-coding region of the *ss* locus located ~7 kb upstream of the transcriptional start (**Fig. 1B and Fig. 3A**). To identify *trans* factors whose binding might be affected by *sin*, we searched for binding motifs neighboring *sin* in bacterial one-hybrid (B1H)(Zhu et al. 2011; Enuameh et al. 2013) and SELEX-seq datasets(Nitta et al. 2015). *sin* lies in a predicted binding site for the zinc finger transcription factor Klumpfuss (Klu), the fly homolog of Wilms' Tumor Suppressor Protein 1 (WT1) (**Fig. 3B, S1A**)(Klein and Campos-Ortega 1997; Yang et al. 1997). This region is perfectly conserved across 21 *Drosophila* species covering 50 million years of evolution, consistent with a critical regulatory role (**Fig. 3C, S1B-C**).

Since *sin* does not affect the core Klu DNA binding motif (MCWCCCMCRC), we predicted that *sin* would not ablate binding, but would rather alter the affinity of Klu for the site. To evaluate the effect of *sin* on Klu

7

binding, we analyzed available SELEX-seq binding data(Nitta et al. 2015). The number of reads containing the Klu binding site with *sin* (CGCCCACAC<u>C</u>) was significantly higher than without *sin* (CGCCCACAC<u>A</u>) (**Fig. 3D**), and thus, Klu binds sequences with *sin* better than those without it. Considering the frequency of 10-mers as a measure of site preference, we found that 506 10-mers (0.10%) have frequencies greater than the Klu site without *sin,* whereas only 366 10-mers (0.07%) have frequencies greater than the Klu site with *sin*. Together, *sin* increases the binding affinity of the Klu site.

Klu lowers the Ss^{ON}/Ss^{OFF} ratio in R7s

As *sin* increases Klu binding affinity and decreases Ss expression frequency, we predicted that Klu acts as a repressor of stochastic *ss* expression in R7s. Consistent with this hypothesis, Klu/WT1 has been shown in other systems to be a transcriptional repressor (Drummond et al. 1992; McDonald et al. 2003; Kaspar et al. 2008). Further, we found that Klu was expressed in all R7s in larval eye imaginal discs (**Fig. 4A-B**)(Wildonger et al. 2005). Since Klu is a repressor, we predicted that increasing Klu levels would cause a decrease in the proportion of Ss^{ON} R7s whereas decreasing or completely ablating Klu would cause an increase in the proportion of Ss^{ON} R7s. Indeed, increasing the levels of Klu in Klu-expressing cells (*klu>klu*), all photoreceptors (*eye>klu*), or specifically in all R7s (*R7>klu*) caused a decrease in the proportion of Ss^{ON} (Rh4) R7s (**Fig. 4C-D**). This decrease in the Ss^{ON}/Ss^{OFF} ratio upon increasing Klu levels mimicked the effect of *sin*, consistent with *sin* increasing the binding affinity for

8

the Klu repressor. Conversely, *klu* null and strong hypomorphic mutants displayed increases in the proportion of Ss^{ON} (Rh4) R7s (**Fig. 4E-F**). Since raising Klu levels decreased the Ss^{ON}/Ss^{OFF} ratio and ablating *klu* increased the ratio, Klu is endogenously expressed at intermediate levels to determine the proportion of Ss-expressing R7s.

Our studies of wild-derived flies revealed significant variation in stochastic Ss expression. We identified *sin*, a single base pair insertion in the ~60 kb *ss* locus that dramatically lowers the Ss^{ON}/Ss^{OFF} ratio by increasing the binding affinity of the transcriptional repressor Klu. This decrease in Ss expression frequency changes the proportions of color-detecting photoreceptors and alters innate color preference in flies.

sin appears to be a relatively new mutation in *D. melanogaster* populations. *sin* is absent among diverse drosophilid species spanning millions of years of divergence (**Fig. S1B-C**) and is segregating at an extremely low frequency among non-admixed African *D. melanogaster* lineages (**Fig. S2A-F**). *sin* likely rose to intermediate frequencies following *D. melanogaster's* colonization of Europe about 10-15 thousand years ago (Li and Stephan 2006). *sin* continues to segregate at intermediate frequencies amongst North American populations (**Fig. S2A-F**), which were established within the last 150 years from mixtures of European and African populations (Bergland et al. 2016). The recent rise in frequency of *sin* suggests that it could be the target of natural selection, perhaps via modulation of innate color preference. We tested this model by

9

assessing patterns of allele frequency differentiation among populations sampled world-wide and also through examination of haplotype homozygostiy surrounding *sin*. We compared these statistics at *sin* to the distribution of statistics calculated from several thousand randomly selected 1-2bp indel polymorphisms that segregate at ~25% in the DGRP. Curiously, *sin* did not deviate from genome-wide patterns (**Fig. S2G-J**) suggesting that it might be selectively neutral in contemporary *D. melanogaster* populations.

It is interesting that Rhodopsin expression varies so significantly in the wild, given the nearly invariant hexagonal lattice of ommatidia in the fly eye. Rhodopsins are G-protein coupled receptors (GPCRs), a class of proteins identified as a source of natural behavioral variation in worms, mice, and voles (Young et al. 1999; Yalcin et al. 2004; Bendesky et al. 2011). Dramatic differences in Rhodopsin expression patterns across insect species (Wernet et al. 2015) suggest that variation in the expression of GPCRs, rather than retinal morphology, may allow rapid evolution in response to environmental changes.

sin increases the binding affinity of a conserved Klu site, suggesting that the site is suboptimal or low affinity for Klu binding. Low affinity sites ensure the timing and specificity of gene expression (Jiang and Levine 1993; Gaudet and Mango 2002; Scardigli et al. 2003; Rowan et al. 2010; Ramos and Barolo 2013; Crocker et al. 2015; Farley et al. 2015; Crocker et al. 2016). Our studies have revealed a critical role for a low affinity binding site in the regulation of stochastic gene expression. The suboptimal Klu site, bound by endogenous levels of Klu, yields the normal 65:35 Ss^{ON}/Ss^{OFF} ratio. Changing the affinity of the site or the

levels of Klu alters the ratio of Ss^{ON}/Ss^{OFF} cells. Together, we conclude that stochastic on/off gene expression is controlled by threshold levels of *trans* factors binding to low affinity sites.

Levels of Klu (analog input) determine the binary on/off ratio of Ss expression (digital output). In contrast, gene regulation is best understood in cases where levels of transcription factors (analog input) regulate the levels of target gene expression (analog output). The on/off nature of Ss expression suggests a cooperative mechanism whereby Klu acts with other factors to repress *ss.* The expression state of *ss* could be determined by the intrinsic variation in Klu levels. In this model, if Klu levels exceed a threshold, *ss* is off, and if Klu levels are below the threshold, *ss* is on. Alternatively, Klu levels could set the threshold for a different gene regulatory mechanism, such as DNA looping or heterochromatin spreading. The requirement of two silencers for repression is consistent with a role for DNA looping in stochastic *ss* expression (Johnston and Desplan 2014). Klu levels could shift the balance between DNA looping states that determine on or off *ss* expression.

Cell fate specification is commonly thought of as a reproducible process whereby cell types uniformly express specific batteries of genes. This reproducibility is often the result of high levels of transcription factors binding to high affinity sites, far exceeding a regulatory threshold, yielding expression of target genes in all cells of a given type. In contrast, the stochastic on/off expression of Ss requires finely tuned levels of regulators binding to low affinity sites. We predict that fine tuning of binding site affinities and transcription factor

11

levels will emerge as a common mechanistic feature that determines the ratio of alternative fates in stochastic systems.

Materials and Methods – (also see Supplemental Materials and Methods)

Drosophila genotypes and stocks

Flies were raised on standard cornmeal-molasses-agar medium and grown at 25°C.

Antibody Staining

Adult retinas and larval eye discs were dissected as described (Hsiao et al. 2012).

Quantification of Expression

Frequency of Rh3 (Ss^{OFF}) and Rh4 (Ss^{ON}) expression in R7s was scored in adults. Six or more retinas were scored for each genotype (N). 100 or more R7s were scored for each retina (n). Frequency was assessed using custom semi-automated software (see below) or manually. Error bars indicate standard deviation (SD).

Image Processing

We employed a custom algorithm to identify the positions of individual R7 photoreceptors within an image of the fly retina. The script that implements our algorithm is available at https://app.assembla.com/spaces/roberts-lab-public/wiki/Fly_Retina_Analysis.

Genome-Wide Association Studies

Genotype data from the DGRP freeze 2 lifted to the dm6/BDGP6 release of the *D. melanogaster* genome was obtained from (ftp://ftp.hgsc.bcm.edu/DGRP/). Phenotypes were calculated for the progeny of crosses of DGRP lines and *Df(3R)Exel6269* flies. To estimate genotypes of these flies from the DGRP data, we simulated each cross. For each SNP or indel variant in the DGRP genotype data, we assigned a new genotype: 1) homozygous reference remains homozygous reference, 2) homozygous alternate maps to homozygous alternate *if* in deficiency region, otherwise heterozygous, and 3) all other genotypes mapped to missing or unknown and not included in subsequent analyses. We performed quantitative trait association analysis using plink2 --linear (version 1.90 beta 25 Mar 2016; PMID:25722852). To reduce the impact of population structure, we included the first 20 principal components of the standardized genetic relationship matrix as covariates (calculated using plink2 --pca). To empirically correct p-values for each site, we performed a max(T) permutation test with 10,000 permutations (mperm option to plink2).

13

CRISPR-mediated mutagenesis

sin was inserted into a lab stock line using CRISPR (Gratz et al. 2013; Port et al. 2014).

T-maze Behavioral Assays

T-maze assays were conducted as described in (Yamaguchi et al. 2010).

Consensus sequence

For the B1H data sets, WebLogo3 was used to generate position weight matrices (PWMs)(Zhu et al. 2011; Enuameh et al. 2013) (**Fig. 3B, S1A**). For the SELEX-SEQ data sets, MEME-ChIP version 4.11.2 was used to generate PWMs (Machanick and Bailey 2011; Nitta et al. 2015) (ENA: ERX606541-ERX606544).

Conservation analysis

The Klu site and neighboring sequences for 21 *Drosophila* species were obtained from the UCSC genome browser. TOMTOM version 4.11.2. was used to generate the conservation PWM (Gupta et al. 2007) (**Fig. 3C, S1B**).

SELEX-seq analysis

SELEX-seq datasets from (Nitta et al. 2015) were obtained from ENA (ERX606541-ERX606544). For read level analysis, we counted the number of reads containing the Klu binding site with *sin*, without *sin*, and neither site (there were no reads with both sites). We performed McNemar's test to assess

14

significance. We computed the frequency of each 10-mer within each dataset using Jellyfish version 2.2.6 (Marcais and Kingsford 2011). Using these counts, we determined the number of 10-mers with frequency greater than that of the Klu binding site with and without *sin.* Frequencies reported are for the combination of all four SELEX datasets.

Population genetic analyses

We estimated allele frequencies from populations sampled world-wide at *sin* and at other 1-2bp indel polymorphisms. Allele frequency estimates based on pooled resequencing of populations sampled in North America and Europe were obtained from (Bergland et al. 2014) and (Kapun et al. 2016). Allele frequencies based on haplotypes (Lack et al. 2016) were also obtained from populations sampled North America, the Caribbean, Europe, and Africa.

References

Alqadah A, Hsieh YW, Xiong R, Chuang CF. 2016. Stochastic left-right neuronal asymmetry in Caenorhabditis elegans. *Philosophical transactions of the Royal Society of London Series B, Biological sciences* **371**.

Bell ML, Earl JB, Britt SG. 2007. Two types of Drosophila R7 photoreceptor cells are arranged randomly: A model for stochastic cell-fate determination. *Journal of Comparative Neurology* **502**: 75-85.

Bendesky A, Tsunozaki M, Rockman MV, Kruglyak L, Bargmann CI. 2011. Catecholamine receptor polymorphisms affect decision-making in C. elegans. *Nature* **472**: 313-318.

Bergland AO, Behrman EL, O'Brien KR, Schmidt PS, Petrov DA. 2014. Genomic evidence of rapid and stable adaptive oscillations over seasonal time scales in Drosophila. *PLoS Genet* **10**: e1004775.

Bergland AO, Tobler R, Gonzalez J, Schmidt P, Petrov D. 2016. Secondary contact and local adaptation contribute to genome-wide patterns of clinal variation in Drosophila melanogaster. *Mol Ecol* **25**: 1157-1174.

Choe KM, Clandinin TR. 2005. Thinking about visual behavior; learning about photoreceptor function. *Current topics in developmental biology* **69**: 187-213.

Chou W-H, Huber A, Bentrop J, Schulz S, Schwab K, Chadwell LV, Paulsen R, Britt SG. 1999. Patterning of the R7 and R8 photoreceptor cells of Drosophila: evidence for induced and default cell-fate specification. *Development* **126**: 607-616.

Chou WH, Hall KJ, Wilson DB, Wideman CL, Townson SM, Chadwell LV, Britt SG. 1996. Identification of a novel Drosophila opsin reveals specific patterning of the R7 and R8 photoreceptor cells. *Neuron* **17**: 1101-1115.

Crocker J, Abe N, Rinaldi L, McGregor AP, Frankel N, Wang S, Alsawadi A, Valenti P, Plaza S, Payre F et al. 2015. Low Affinity Binding Site Clusters Confer Hox Specificity and Regulatory Robustness. *Cell* **160**: 191-203.

Crocker J, Noon EP, Stern DL. 2016. The Soft Touch: Low-Affinity Transcription Factor Binding Sites in Development and Evolution. *Current topics in developmental biology* **117**: 455-469.

Drummond IA, Madden SL, Rohwer-Nutter P, Bell GI, Sukhatme VP, Rauscher FJ, 3rd. 1992. Repression of the insulin-like growth factor II gene by the Wilms tumor suppressor WT1. *Science* **257**: 674-678.

Enuameh MS, Asriyan Y, Richards A, Christensen RG, Hall VL, Kazemian M, Zhu C, Pham H, Cheng Q, Blatti C et al. 2013. Global analysis of Drosophila Cys(2)-His(2) zinc finger proteins reveals a multitude of novel recognition motifs and binding determinants. *Genome Res* **23**: 928-940.

Farley EK, Olson KM, Zhang W, Brandt AJ, Rokhsar DS, Levine MS. 2015. Suboptimization of developmental enhancers. *Science* **350**: 325-328.

Feiler R, Bjornson R, Kirschfeld K, Mismer D, Rubin GM, Smith DP, Socolich M, Zuker CS. 1992. Ectopic expression of ultraviolet-rhodopsins in the blue

photoreceptor cells of Drosophila: visual physiology and photochemistry of transgenic animals. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **12**: 3862-3868.

Franceschini N, Kirschfeld K, Minke B. 1981. Fluorescence of Photoreceptor Cells Observed in vivo. *Science* **213**: 1264-1267.

Gaudet J, Mango SE. 2002. Regulation of organogenesis by the Caenorhabditis elegans FoxA protein PHA-4. *Science* **295**: 821-825.

Gratz SJ, Wildonger J, Harrison MM, O'Connor-Giles KM. 2013. CRISPR/Cas9-mediated genome engineering and the promise of designer flies on demand. *Fly* **7**: 249.

Gupta S, Stamatoyannopoulos JA, Bailey TL, Noble WS. 2007. Quantifying similarity between motifs. *Genome biology* **8**: R24.

Heisenberg M, Wolf R. 1984. *Vision in Drosophila*: (Springer-Verlag, Berlin).

Hofer H, Carroll J, Neitz J, Neitz M, Williams DR. 2005. Organization of the human trichromatic cone mosaic. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **25**: 9669-9679.

Hsiao HY, Johnston RJ, Jukam D, Vasiliauskas D, Desplan C, Rister J. 2012. Dissection and immunohistochemistry of larval, pupal and adult Drosophila retinas. *Journal of visualized experiments : JoVE*: e4347.

Hsiao HY, Jukam D, Johnston R, Desplan C. 2013. The neuronal transcription factor erect wing regulates specification and maintenance of Drosophila R8 photoreceptor subtypes. *Dev Biol* **381**: 482-490.

Huber A, Schulz S, Bentrop J, Groell C, Wolfrum U, Paulsen R. 1997. Molecular cloning of Drosophila Rh6 rhodopsin: the visual pigment of a subset of R8 photoreceptor cells. *FEBS Lett* **406**: 6-10.

Jiang J, Levine M. 1993. Binding affinities and cooperative interactions with bHLH activators delimit threshold responses to the dorsal gradient morphogen. *Cell* **72**: 741-752.

Johnston RJ. 2013. Lessons about terminal differentiation from the specification of color-detecting photoreceptors in the Drosophila retina. *Annals of the New York Academy of Sciences* **1293**: 33-44.

Johnston RJ, Desplan C. 2014. Interchromosomal Communication Coordinates Intrinsically Stochastic Expression Between Alleles. *Science* **343**: 661-665.

Johnston RJ, Jr., Desplan C. 2010. Stochastic mechanisms of cell fate specification that yield random or robust outcomes. *Annual review of cell and developmental biology* **26**: 689-719.

Johnston RJ, Jr., Otake Y, Sood P, Vogt N, Behnia R, Vasiliauskas D, McDonald E, Xie B, Koenig S, Wolf R et al. 2011. Interlocked feedforward loops control cell-type-specific Rhodopsin expression in the Drosophila eye. *Cell* **145**: 956-968.

Jukam D, Desplan C. 2011. Binary regulation of Hippo pathway by Merlin/NF2, Kibra, Lgl, and Melted specifies and maintains postmitotic neuronal fate. *Developmental cell* **21**: 874-887.

Jukam D, Viets K, Anderson C, Zhou C, DeFord P, Yan J, Cao J, Johnston RJ, Jr. 2016. The insulator protein BEAF-32 is required for Hippo pathway

17

activity in the terminal differentiation of neuronal subtypes. *Development* **143**: 2389-2397.

Jukam D, Xie B, Rister J, Terrell D, Charlton-Perkins M, Pistillo D, Gebelein B, Desplan C, Cook T. 2013. Opposite feedbacks in the Hippo pathway for growth control and neural fate. *Science* **342**: 1238016.

Kapun M, Fabian DK, Goudet J, Flatt T. 2016. Genomic Evidence for Adaptive Inversion Clines in Drosophila melanogaster. *Molecular biology and evolution* **33**: 1317-1336.

Kaspar M, Schneider M, Chia W, Klein T. 2008. Klumpfuss is involved in the determination of sensory organ precursors in< i> Drosophila</i>. *Developmental biology* **324**: 177-191.

Klein T, Campos-Ortega JA. 1997. klumpfuss, a Drosophila gene encoding a member of the EGR family of transcription factors, is involved in bristle and leg development. *Development* **124**: 3123-3134.

Lack JB, Cardeno CM, Crepeau MW, Taylor W, Corbett-Detig RB, Stevens KA, Langley CH, Pool JE. 2015. The Drosophila genome nexus: a population genomic resource of 623 Drosophila melanogaster genomes, including 197 from a single ancestral range population. *Genetics* **199**: 1229-1241.

Lack JB, Lange JD, Tang AD, Corbett-Detig RB, Pool JE. 2016. A Thousand Fly Genomes: An Expanded Drosophila Genome Nexus. *Molecular biology and evolution* **33**: 3308-3313.

Li H, Stephan W. 2006. Inferring the demographic history and rate of adaptive substitution in Drosophila. *PLoS Genet* **2**: e166.

Machanick P, Bailey TL. 2011. MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics* **27**: 1696-1697.

Mackay TF, Richards S, Stone EA, Barbadilla A, Ayroles JF, Zhu D, Casillas S, Han Y, Magwire MM, Cridland JM. 2012. The Drosophila melanogaster genetic reference panel. *Nature* **482**: 173-178.

Magklara A, Lomvardas S. 2013. Stochastic gene expression in mammals: lessons from olfaction. *Trends Cell Biol* **23**: 449-456.

Marcais G, Kingsford C. 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**: 764-770.

McDonald JA, Fujioka M, Odden JP, Jaynes JB, Doe CQ. 2003. Specification of motoneuron fate in Drosophila: integration of positive and negative transcription factor inputs by a minimal eve enhancer. *Journal of neurobiology* **57**: 193-203.

McEwen RS. 1918. The reactions to light and to gravity in Drosophila and its mutants. *J Exp Zool* **25**: 49–106.

Mikeladze-Dvali T, Wernet MF, Pistillo D, Mazzoni EO, Teleman AA, Chen YW, Cohen S, Desplan C. 2005. The growth regulators warts/lats and melted interact in a bistable loop to specify opposite fates in Drosophila R8 photoreceptors. *Cell* **122**: 775-787.

Montell C, Jones K, Zuker C, Rubin G. 1987. A second opsin gene expressed in the ultraviolet-sensitive R7 photoreceptor cells of Drosophila melanogaster. *The Journal of neuroscience* **7**: 1558-1566.

Nitta KR, Jolma A, Yin Y, Morgunova E, Kivioja T, Akhtar J, Hens K, Toivonen J, Deplancke B, Furlong EE. 2015. Conservation of transcription factor binding specificities across 600 million years of bilateria evolution. *Elife* **4**: e04837.

Port F, Chen H-M, Lee T, Bullock SL. 2014. An optimized CRISPR/Cas toolbox for efficient germline and somatic genome engineering in Drosophila. *bioRxiv*.

Ramos AI, Barolo S. 2013. Low-affinity transcription factor binding sites shape morphogen responses and enhancer evolution. *Philosophical transactions of the Royal Society of London Series B, Biological sciences* **368**: 20130018.

Ressler KJ, Sullivan SL, Buck LB. 1993. A zonal organization of odorant receptor gene expression in the olfactory epithelium. *Cell* **73**: 597-609.

Roorda A, Williams DR. 1999. The arrangement of the three cone classes in the living human eye. *Nature* **397**: 520-522.

Rowan S, Siggers T, Lachke SA, Yue Y, Bulyk ML, Maas RL. 2010. Precise temporal control of the eye regulatory gene Pax6 via enhancer-binding site affinity. *Genes Dev* **24**: 980-985.

Salcedo E, Huber A, Henrich S, Chadwell LV, Chou WH, Paulsen R, Britt SG. 1999. Blue- and green-absorbing visual pigments of Drosophila: ectopic expression and physiological characterization of the R8 photoreceptor cell-specific Rh5 and Rh6 rhodopsins. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **19**: 10716-10726.

Scardigli R, Bäumer N, Gruss P, Guillemot F, Le Roux I. 2003. Direct and concentration-dependent regulation of the proneural gene Neurogenin2 by Pax6. *Development* **130**: 3269-3281.

Thanawala Shivani U, Rister J, Goldberg Gregory W, Zuskov A, Olesnicky Eugenia C, Flowers Jonathan M, Jukam D, Purugganan Michael D, Gavis Elizabeth R, Desplan C et al. 2013. Regional Modulation of a Stochastically Expressed Factor Determines Photoreceptor Subtypes in the Drosophila Retina. *Developmental cell* **25**: 93-105.

Troemel ER, Sagasti A, Bargmann CI. 1999. Lateral Signaling Mediated by Axon Contact and Calcium Entry Regulates Asymmetric Odorant Receptor Expression in C. elegans. *Cell* **99**: 387-398.

Tully T, Quinn WG. 1985. Classical conditioning and retention in normal and mutant Drosophila melanogaster. *Journal of comparative physiology A, Sensory, neural, and behavioral physiology* **157**: 263-277.

Viets K, Eldred KC, Johnston RJ, Jr. 2016. Mechanisms of Photoreceptor Patterning in Vertebrates and Invertebrates. *Trends Genet* **32**: 638-659.

Wernet MF, Mazzoni EO, Çelik A, Duncan DM, Duncan I, Desplan C. 2006. Stochastic spineless expression creates the retinal mosaic for colour vision. *Nature* **440**: 174-180.

Wernet MF, Perry MW, Desplan C. 2015. The evolutionary diversity of insect retinal mosaics: common design principles and emerging molecular logic. *Trends Genet* **31**: 316-328.

Wildonger J, Sosinsky A, Honig B, Mann RS. 2005. Lozenge directly activates argos and klumpfuss to regulate programmed cell death. *Genes & Development* **19**: 1034-1039.

Yalcin B, Willis-Owen SA, Fullerton J, Meesaq A, Deacon RM, Rawlins JN, Copley RR, Morris AP, Flint J, Mott R. 2004. Genetic dissection of a behavioral quantitative trait locus shows that Rgs2 modulates anxiety in mice. *Nature genetics* **36**: 1197-1202.

Yamaguchi S, Desplan C, Heisenberg M. 2010. Contribution of photoreceptor subtypes to spectral wavelength preference in Drosophila. *Proceedings of the National Academy of Sciences of the United States of America* **107**: 5634-5639.

Yang X, Bahri S, Klein T, Chia W. 1997. Klumpfuss, a putative Drosophila zinc finger transcription factor, acts to differentiate between the identities of two secondary precursor cells within one neuroblast lineage. *Genes & Development* **11**: 1396-1408.

Young LJ, Nilsen R, Waymire KG, MacGregor GR, Insel TR. 1999. Increased affiliative response to vasopressin in mice expressing the V1a receptor from a monogamous vole. *Nature* **400**: 766-768.

Zhu LJ, Christensen RG, Kazemian M, Hull CJ, Enuameh MS, Basciotta MD, Brasefield JA, Zhu C, Asriyan Y, Lapointe DS. 2011. FlyFactorSurvey: a database of Drosophila transcription factor binding specificities determined using the bacterial one-hybrid system. *Nucleic acids research* **39**: D111-D117.

Zuker CS, Montell C, Jones K, Laverty T, Rubin GM. 1987. A rhodopsin gene expressed in photoreceptor cell R7 of the Drosophila eye: homologies with other signal-transducing molecules. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **7**: 1550-1557.

Figure Legends

Figure 1. A naturally-occurring single base insertion in the *ss* locus (*sin*) lowers the ratio of Ss^{ON} to Ss^{OFF} R7s

A. R7 and R8 subtypes are determined by the on/off expression of Spineless (Ss).

(Left) The absence of Ss allows Rh3 expression in pale R7s and Rh5 expression in pale R8s.

(Right) Expression of Ss induces Rh4 expression in yellow R7s and Rh6 expression in yellow R8s.

The signal by which Spineless mediates Rh5 vs. Rh6 expression in R8s is currently unknown.

B. Schematic of the *ss* locus. Green dashed rectangle indicates *R7/R8 enhancer*; red dashed rectangles indicate *silencer 1* and *silencer 2*; blue line indicates Klu binding site; blue arrow indicates *ss insertion/sin;* gray ovals represent untranslated exons; yellow ovals represent translated exons; black boxes indicate neighboring genes; arrows indicate transcriptional starts.

C. Image of a whole mount fly retina.

(Left) Stochastic distribution of R7s expressing Rh3 (Ss^{OFF}) or Rh4 (Ss^{ON}).

(Right) An automated counting system identified and counted Rh3- and Rh4-expressing R7s.

21

D. Crossing scheme: Wild-derived DGRP flies were crossed with *ss* deficiency flies, yielding progeny that were hemizygous at the *ss* locus. Orange lines indicate genetic variants; blue line indicates *sin* in *ss.*

E. Representative images from progeny in (D) with low (left; DGRP-397) and high (right; DGRP-229) proportions of Ss^{ON} (Rh4) R7s.

F. Ss^{ON} proportion varied across DGRP fly lines. *sin* was enriched in lines with a low proportion of Ss^{ON} R7s. Each bar represents a single DGRP line, and bars are arranged in rank order. Light blue bars indicate hemizygous *sin*. Dark blue bars indicate hemizygous *sin* or hemizygous *no sin* (original DGRP line was heterozygous *sin/no sin*). Gray bars indicate hemizygous *no sin*.

G. GWAS identified *sin* as a genetic variant associated with Ss expression. Manhattan plot of the genetic variant p-values. Genetic variants above the red line (Bonferroni correction) are considered significant. Arrow indicates *sin*.

H. *sin* was enriched in lines with a low proportion of Ss^{ON} R7s. Violin plot of DGRP lines with and without *sin*.

I-K. Flies with *sin* displayed a lower proportion of Ss^{ON} R7s compared to flies without *sin.* AL indicates African lines; LL indicates laboratory lines; CL indicates lines in which *sin* was inserted with CRISPR. **** indicates p<0.0001; *** indicates p<0.001. Error bars indicate standard deviation (SD).

Figure 2. Wild-derived flies with *sin* display a shift in innate color preference from green to blue

22

A. Schematic of T-maze apparatus. Red dots represent flies. PI = Preference Index. Positive PI indicates preference for green light; negative PI indicates preference for blue light. N_G: number of flies on green side; N_B: number of flies on blue side.

B-C. Flies with *sin* preferred blue light compared to flies without *sin* that preferred green light. Color of bar indicates genotype of DGRP line. Light blue bars indicate homozygous *sin*. Dark blue bars indicate heterozygous *sin* / *no sin.* Gray bars indicate homozygous *no sin*. Error bars indicate standard error of the mean (SEM).

B. PIs for individual DGRP lines with and without *sin.*

C. Averages of PIs for DGRP lines with and without *sin.*

Figure 3. *sin* increases the binding affinity for the transcription factor Klumpfuss

A. *sin* is a single base pair insertion of a C at Chr. 3R: 16,410,775 (release 6). Underline indicates *sin.*

B. *sin* affects a predicted binding site for the transcription factor Klumpfuss (Klu). Binding site predicted from B1H.

C. The Klu site is perfectly conserved across 21 species of *Drosophila* covering 50 million years of evolution. Conservation logo of the Klu site and neighboring sequence in the *ss* locus. Height of bases indicates degree of conservation.

D. *sin* increased Klu binding affinity *in vitro.* Quantification of the number of reads for the Klu site with and without *sin* in four cycles of SELEX-seq. * indicates $p < 0.05$ in cycle 1; **** indicates $p < 0.0001$ in cycles 2-4.

Figure 4. Levels of Klu determine the ratio of Ss^{ON}/Ss^{OFF} R7s

A. Klu was expressed in the developing larval eye disc. MF indicates morphogenetic furrow.

B. Klu was expressed in all R7s in the developing larval eye disc. Red indicates R7s marked by *pm181>Gal4, UAS>mcd8GFP.*

C-D. Increasing Klu levels decreased the proportion of Ss^{ON} R7s. In C, representative image of R7s expressing Rh3 (Ss^{OFF}) or Rh4 (Ss^{ON}) in *eye>klu* flies. **** indicates $p < 0.0001$. Error bars indicate standard deviation (SD).

E-F. *klu* loss of function mutants displayed increases in the proportion of Ss^{ON} R7s. In E, representative image of R7s expressing Rh3 (Ss^{OFF}) or Rh4 (Ss^{ON}) in *klu* hypomorphic flies. **** indicates $p < 0.0001$; *** indicates $p < 0.001$. Error bars indicate standard deviation (SD).
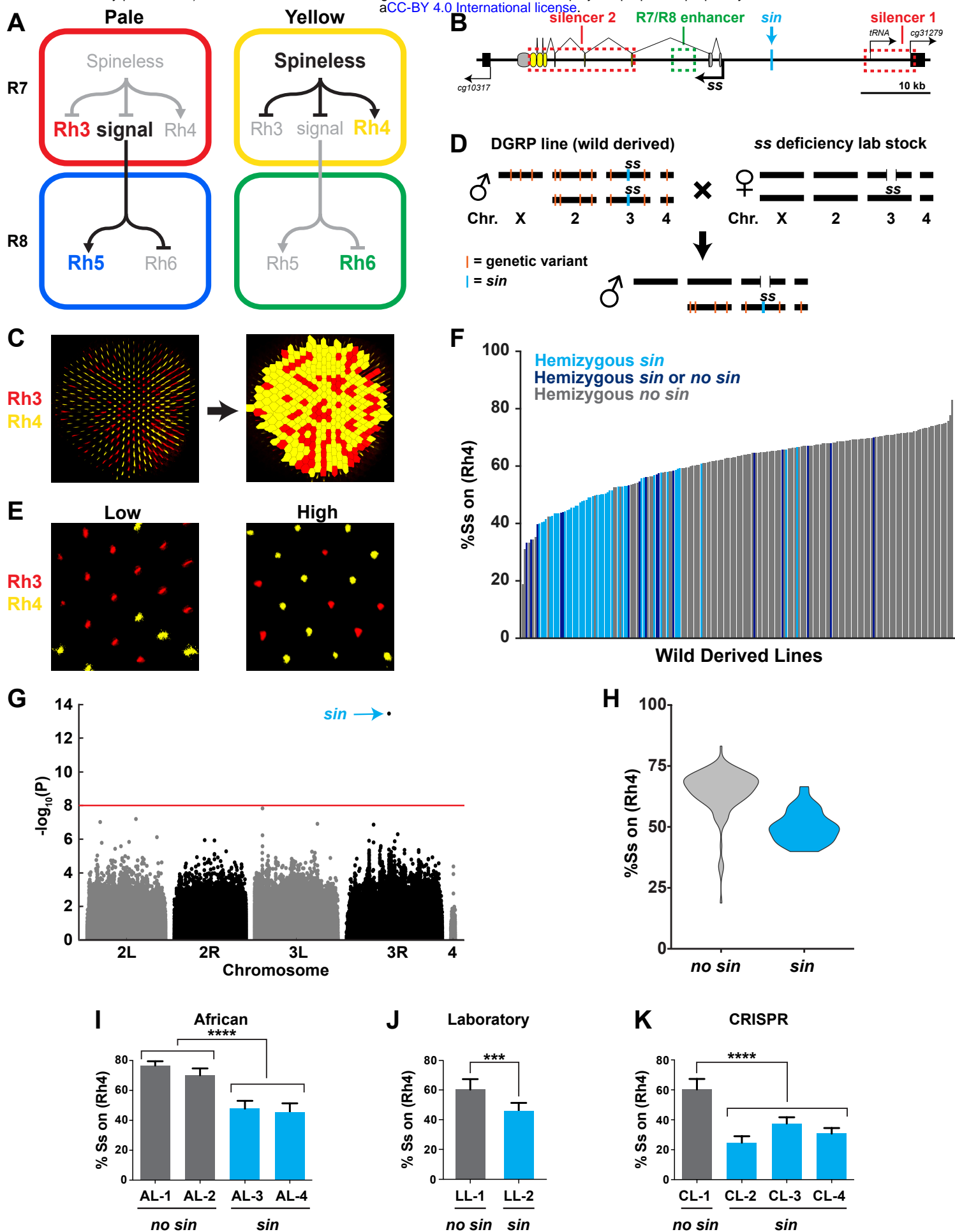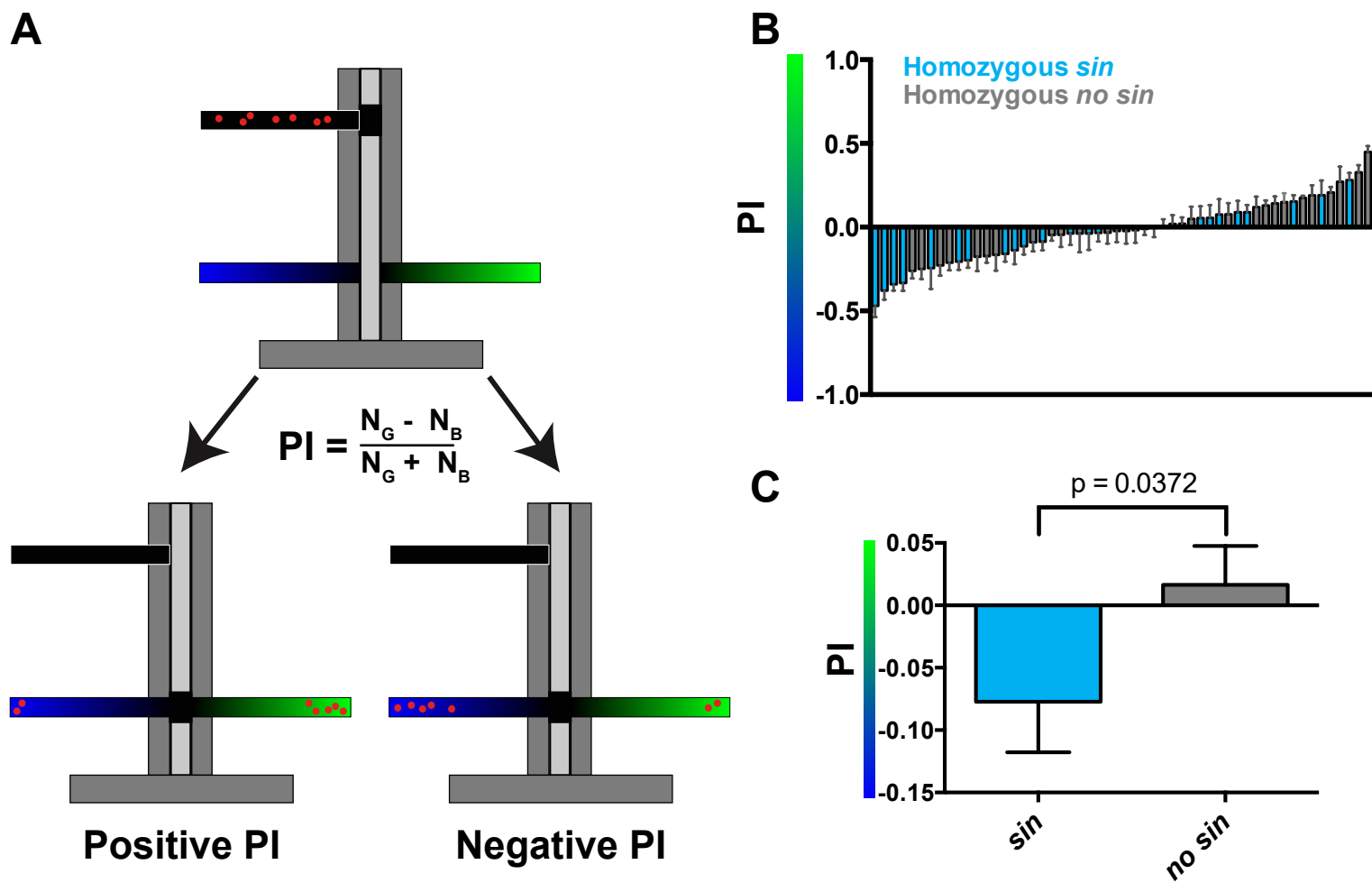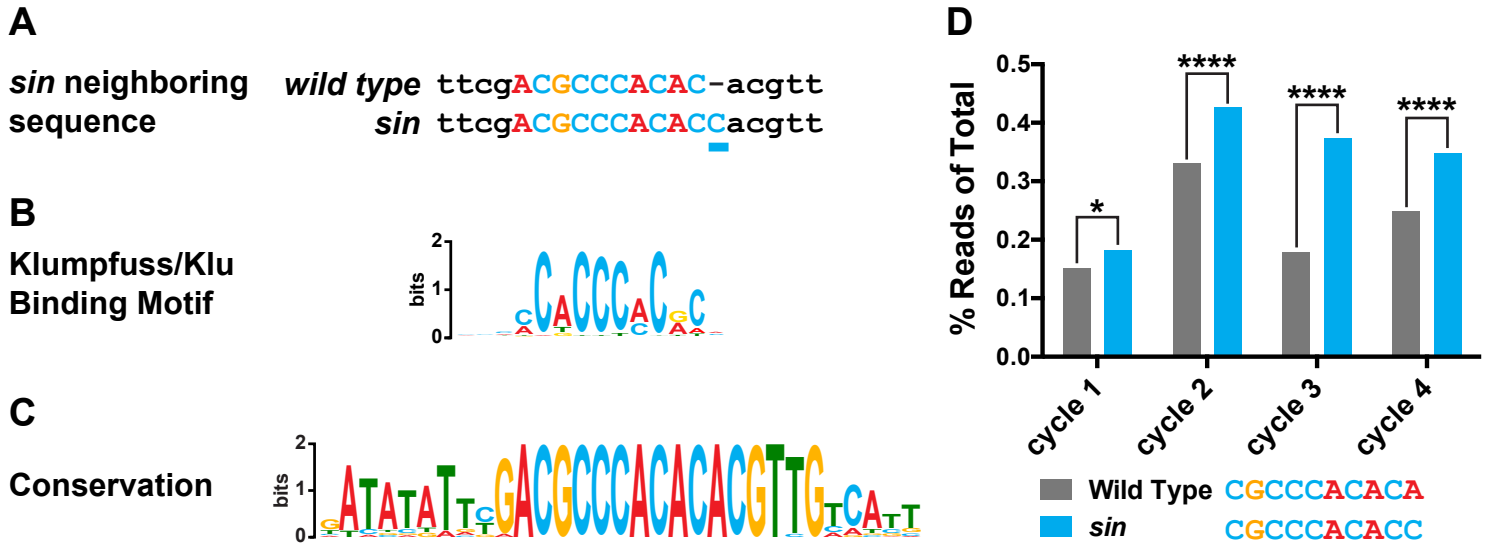
Figure 1

A

$$PI = \frac{N_G - N_B}{N_G + N_B}$$

Positive PI

Negative PI

B

Homozygous *sin*
Homozygous *no sin*

C

p = 0.0372

sin   no sin

Figure 2

A

***sin* neighboring sequence**

wild type ttcgACGCCCACAC-acgtt
sin ttcgACGCCCACACCacgtt

B

Klumpfuss/Klu Binding Motif

C

Conservation

D

Figure 3

Figure 4