# Dissecting *cis*-regulatory control of quantitative trait variation in a plant stem cell circuit

Xingang Wang[1], Lyndsey Aguirre[1,2], Daniel Rodríguez-Leal[1,4], Anat Hendelman[1], Matthias Benoit[1,3], and Zachary B. Lippman[1,2,3*]

[1] Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA
[2] School of Biological Sciences, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA
[3] Howard Hughes Medical Institute, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA
[4] Present address: Inari Agriculture, Cambridge, MA, USA

* To whom correspondence may be addressed:

lippman@cshl.edu

***Cis*-regulatory mutations underlie important crop domestication and improvement traits[1,2]. However, limited allelic diversity has hindered functional dissection of the large number of *cis*-regulatory elements and their potential interactions, thereby precluding a deeper understanding of how *cis*-regulatory variation impacts traits quantitatively. Here, we engineered over 60 promoter alleles in two tomato fruit size genes[3,4] to characterize *cis*-regulatory sequences and study their functional relationships. We found that targeted mutations in conserved promoter sequences of *SlCLV3,* a repressor of stem cell proliferation[5,6], have a weak impact on fruit locule number. Pairwise combinations of these mutations mildly enhance this phenotype, revealing additive and synergistic relationships between conserved regions, and further suggesting even higher-order *cis*-regulatory interactions within the *SlCLV3* promoter. In contrast, *SlWUS*, a positive regulator of stem cell proliferation repressed by *SlCLV3*[5,6]*,* is more tolerant to promoter perturbations. Our results show that complex interplay among *cis*-regulatory variants can shape quantitative variation, and suggest that empirical dissections of this hidden complexity can guide promoter engineering to predictably modify crop traits.**

*Cis*-regulatory DNA determines patterns and levels of gene expression, and decoding this regulatory information is essential in understanding how genotypes translate to phenotypes. The vast *cis*-regulatory space surrounding genes makes it challenging to identify functional sequences[7]. Recent studies in diverse plant species have predicted the genome-wide presence of *cis*-regulatory elements (CREs) using sequence conservation, transcription factor binding, chromatin accessibility, and other molecular and computational approaches[8–15]. However, empirical characterization of whether and to what extent these sequences regulate phenotypes are lagging far behind.

The identification of rare, natural mutations contributing to crop domestication and improvement has illuminated the importance of *cis*-regulatory regions in controlling quantitative trait variation[1,2,16]. Emerging pan-genomes have exposed expansive *cis*-regulatory variation, including simple variants (e.g. SNPs, indels) and more complex structural variants (SVs), which are often associated with modified expression and phenotypes[17–21]. However, identifying causative mutations is challenging, as variants with subtle effects are difficult to resolve and multiple mutations within and between *cis*-regulatory regions could be acting together to

influence phenotypes[22–24]. Thus, the limited number of characterized alleles has been insufficient to dissect the functional components of a gene's *cis*-regulatory space, leaving it unclear why specific genetic perturbations result in specific quantitative phenotypic outputs. Resolving these relationships is key for the precise design and engineering of *cis*-regulatory alleles with predictable effects on crop improvement[16,25]. Here, we use genome editing to finely dissect *cis*-regulatory control of quantitative trait variation in two genes controlling stem cell proliferation and fruit size in tomato.

The *CLAVATA3* (*CLV3*) gene encodes a conserved small signaling peptide that inhibits stem cell proliferation in the shoot apical meristem in many plants[5,6,26]. Similar to other species[27–30], loss of tomato (*Solanum lycopersicum,* denoted with '*Sl*' prefix) *SlCLV3* results in enlarged meristems that cause fasciated phenotypes, including many more seed compartments (locules) in fruits compared to wild type plants (WT)[4,25]. We previously developed a CRISPR-Cas9 multiplex mutagenesis drive system to engineer quantitative trait loci (QTLs) for crop improvement[25]. Using this tool, we generated 15 *SlCLV3* promoter (*slclv3$^{pro}$*) alleles, which resulted in a range of fruit locule number variation[25]. However, the limited number of alleles, each having multiple mutations in a 1.7 kb target region, precluded association of specific promoter sequences with quantitative phenotypic changes. Hence, to increase mapping resolution, we used the same CRISPR-Cas9 drive system with eight gRNAs to generate 14 new *slclv3$^{pro}$* alleles (**Fig. 1a**). The resulting series of 30 alleles, including the natural QTL inversion allele *fasciated* (*fas*) and a null allele that eliminates 7.3 kb of *SlCLV3* promoter and coding sequence (*slclv3$^{pro\text{-}29}$*), contained various types of mutations, such as large deletions, inversions, and small indels across the target region. To simplify their visualization, we encoded each promoter allele using heatmap representations of sequence modifications in sequential 20 bp windows (**Fig. 1b, c**). Arranging the *slclv3$^{pro}$* alleles by phenotypic strength revealed a continuum of locule number variation, and trends in associations between *cis*-regulatory mutations and phenotypes. 14 of the alleles had weak increases in locule number (*slclv3$^{pro\text{-}4}$* to *slclv3$^{pro\text{-}17}$*; ~1-3 more locules than WT) and were associated primarily with deletions that disrupted the proximal half of the target region (**Fig. 1d and Extended Data Fig. 1a,b**). In contrast, most of the *slclv3$^{pro}$* alleles with locule number increases greater than *fas* (mean 6.1 locules) disrupted the distal half of the target region, and often contained mutations in both proximal and distal regions. Two alleles with the strongest effects on locule number removed most (*slclv3$^{pro\text{-}27}$*) or all of the target region (*slclv3$^{pro\text{-}28}$*) and

nearly matched the effect of the null allele (*slclv3*$^{pro\text{-}29}$, mean 15.8 locules). Though there were exceptions to these trends, likely due to multiple and different combinations of mutations across all promoter alleles, the expanded *slclv3*$^{pro}$ allelic series indicated that multiple sequences throughout the *SlCLV3* promoter are important for its function, with a more prominent role for sequences in the distal region.

However, this expanded allelic diversity was still insufficient to relate specific *cis*-regulatory regions to quantitative phenotypic effects. Sequence conservation could indicate *cis*-regulatory function[11,31]. We searched for conserved non-coding sequences (CNSs) by aligning the *SlCLV3* promoter sequence with corresponding regions from the related Solanaceae species potato (*Solanum tuberosum*), pepper (*Capsicum annuum*), and groundcherry (*Physalis grisea*), representing 25 million years of evolution. This analysis (see Methods) identified three deeply conserved CNS regions (designated R1, R3, R4) (**Fig. 2a**), and an additional CNS region (R2) is shared only between tomato and potato. We also identified dozens of predicted transcription factor binding sites (TFBSs) throughout the entire *SlCLV3* promoter, including many in the CNS regions; however, the abundance of these sequences makes precise and systematic functional characterizations impracticable (**Fig. 2a**).

We therefore used the available gRNA recognition sites to design four new CRISPR-Cas9 constructs, each having three or four gRNAs to remove all or large portions of each CNS region. We generated 16 alleles in total, with at least three alleles for each region (**Fig. 2b-e**). Five deletion alleles in R1 ranged in size from 3 bp to 226 bp, and the two smallest deletions (3 bp and 63 bp deletion) had no effect on locule number, suggesting the underlying sequences are not critical for promoter function. Notably, only the three largest R1 deletion alleles, from allele *slclv3*$^{pro\text{-}R1\text{-}3}$ (73 bp deletion) to *slclv3*$^{pro\text{-}R1\text{-}5}$ (226 bp deletion, removing the entire targeted region and most of the R1 CNS), caused weak increases in locule number compared to WT (**Fig. 2b**). Similarly, none of the other 11 deletion alleles in the other CNS regions had a substantial effect on locule number, and only the two largest R4 deletion alleles (*slclv3*$^{pro\text{-}R4\text{-}4}$: 91 bp deletion and *slclv3*$^{pro\text{-}R4\text{-}5}$: 340 bp deletion) weakly increased locule number, similar to the largest R1 deletions. These observations suggest that disruption of multiple sequences within a conserved region is likely required to translate into a phenotypic effect. The absence of phenotypes from R2 alleles, including *slclv3*$^{pro\text{-}R2\text{-}3}$ that disrupted the entire R2 region, is consistent with this CNS region being less conserved throughout Solanaceae, and the largest R3 allele removed only half of the

target region, possibly leaving functional sequences intact (**Figure 2c-e and Supplemental Fig. 2a-d**). Most of the CNS alleles impacted at least one TFBS, and many of the larger deletions removed many of them. However, loss of multiple TFBSs did not always result in a phenotype (e.g. *slclv3*$^{pro\text{-}R1\text{-}5}$ vs. *slclv3*$^{pro\text{-}R2\text{-}3}$), suggesting we could not use these sites as predictors of phenotypic effects. We also tested if changes in *SlCLV3* expression were associated with phenotypes, and only subtle expression differences were detected in these alleles. Consistent with previous observations[25], there were no strong correlations between altered expression and phenotypic effects (**Extended Data Fig. 2e-i**). Together, these results show that CNSs contain functional sequences, and that the R1 and R4 CNS regions are important for *SlCLV3* promoter function, but only partially contribute to its activity.

The weak phenotypes of individual R1 and R4 deletion alleles compared to the alleles that removed both proximal and distal promoter sequences (**Fig. 1b-d** and **Fig. 2b-e**) suggested genetic interactions between conserved regions. To explore these relationships, we devised two strategies to create alleles with combinations of mutations in two different CNS regions (**Extended Data Fig. 3a,b**), since their close physical distances prevented combining them by recombination. Our approaches preserve an existing mutation in one CNS region and then introduce a new mutation in a second CNS region, which avoids altering sequences between two targeted regions. In the first approach, trans-targeting, we crossed plants homozygous for individual mutations that also carry their respective CRISPR-Cas9 transgenes to allow for reciprocal targeting of the inherited wild type CNSs. Alleles with mutations in two CNSs were then identified in F1 plants, and homozygous mutants were recovered in F2 populations (**Extended Data Fig. 3a**). Our second approach used sequential editing, in which transgene-free homozygous mutants having a deletion in one region were transformed with a CRISPR-Cas9 construct targeting a second region (**Extended Data Fig. 3b**). Since only the largest deletion alleles in our individual CNS targeting resulted in phenotypes, we focused on isolating new alleles with large perturbations in combinations of CNS regions. Applying both approaches, we obtained a total of 13 pairwise combined mutations in R1-R4, R1-R2, and R2-R4, with at least two alleles for each combination of targeted CNS regions.

Nine new alleles with mutations in both R1 and R4 all caused a greater increase in locule number compared to the strongest allele from each individual region (**Fig. 3a and Extended Data Fig. 3c**), with different R1-R4 combined alleles enhancing the phenotype to varying

degrees (**Fig. 3a**, e.g. $R1^{5}$+$R4^{c}$ vs. $R1^{b}$+$R4^{5}$). This result prompted us to test if the enhanced phenotypes were the sum of effects from individual R1 and R4 mutations, or if they exceeded them (i.e. additivity vs. synergism). Since the newly induced mutations in the combined alleles were different from the original individual mutations, we performed stringent tests of additivity or synergism by using the sum of the effects from the strongest individual alleles in each region. Our statistical analyses (see Methods) showed that four of the nine R1-R4 combined alleles had synergistic effects (**Fig. 3a,b and Extended Data Fig. 3c,d**), and one ($R1^{5}$+$R4^{d}$) was synergistic in one experiment and additive in another, possibly due to environmental influence. The remaining four combined alleles had additive effects, and interestingly three of them had small indels in either R1 or R4 (**Extended Data Fig. 3c,d**), which overlapped with indels that showed no effect on their own (**Fig. 2b-e**). This suggested that some mutations exhibit phenotypic effects only in the presence of other mutations, reflecting redundant relationships between the underlying sequences. A similar relationship was evident between some mutations in R2, which have no effect on their own, and mutations in other regions (**Fig. 2c**). For example, the combined $R1^{5}$+$R2^{a}$ allele, which inherited the original R1 mutation and a partial R2 deletion, resulted in similar locule numbers as the single R1 mutation (**Fig. 3c**), whereas an allele that removed R1 and R2 together (*R1+R2*) resulted in a much stronger phenotype. These observations suggest that the R2 CNS functions redundantly with the R1 CNS; however, this allele disrupted an additional 40 bp near R1 relative to the original single mutation, which could be contributing to enhancement (**Fig. 3c and Extended Data Fig. 3e**). Such redundant effects may be specific, as combining R2 and R4 mutations did not increase locule number compared to R4 alleles alone, though R2 was not entirely deleted in the R2-R4 combined alleles (**Fig. 3d**). Together, these results show that additive, synergistic, and redundant relationships among conserved sequences all contribute to *SlCLV3* promoter function (**Fig. 3e**). Notably, the strongest combined alleles still showed only moderate phenotypic effects, indicating even higher-order interactions underlie the wide range of quantitative variation from the *SlCLV3* promoter (**Fig. 1b-d**).

CLV3 functions in a deeply conserved negative feedback relationship with the homeodomain transcription factor WUSCHEL (WUS), which promotes stem-cell proliferation (**Fig. 4a**)[5,6]. A weak gain-of-function *cis*-regulatory allele (*lc*) that disrupts sequences downstream of tomato *WUS* (*SlWUS*) underlies a locule number QTL with a similar effect as weak $slclv3^{pro}$ alleles[3,25]. We therefore tested whether similar *cis*-regulatory complexity controls

*SlWUS*. Null mutations of *WUS* in *Arabidopsis thaliana* cause premature termination of the primary shoot meristem during embryogenesis, and produce axillary meristems that cycle between reinitiation and termination of vegetative and floral meristems after a few organs have formed. In contrast, hypomorphic *wus* alleles can form several leaves before primary meristem termination, and axillary meristems give rise to shoots and normal flowers[32–34]. As there are no known loss-of-function mutations in *SlWUS*, we generated two frame-shift alleles by targeting its coding sequence with two gRNAs (**Fig. 4b**). Similar to Arabidopsis, homozygous null *slwus* mutant seedlings failed to maintain the shoot apical meristem, which terminated after producing 2-3 leaves (**Fig. 4c**). Reinitiated meristems would then develop and produce a leaf before terminating, resulting in stunted bushy plants that never produced shoots or transitioned to reproductive growth (**Fig 4c,d**). To test potential quantitative effects from *SlWUS* promoter alleles, we performed CRISPR-Cas9 multiplex mutagenesis on a 2.6 kb target region that included four CNS regions (out of five in the *SlWUS* promoter), and generated eight diverse alleles having mostly large deletions that removed one or more CNSs and also intervening sequences (**Fig. 4e,f**). The most severe allele (*slwus*$^{pro\text{-}8}$, 1.9 kb deletion), which eliminated three CNSs and had a rearrangement in the proximal CNSs that could not be resolved, was similar to the null coding sequence mutants (**Extended Data Fig. 4a**). Interestingly, all other *slwus*$^{pro}$ mutants appeared normal, including flower and fruit development, although they occasionally produced extra cotyledons. We asked if these promoter alleles caused weak effects on locule number; however, all were similar to WT, with the exception of allele *slwus*$^{pro\text{-}6}$. This allele contained a 223 bp insertion and a 554 bp inversion and caused a subtle increase in locule number similar to *lc* (**Fig. 4f, g and Extended Data Fig. 4b**). These results suggest that the *SlWUS* promoter is more tolerant to perturbations than the *SlCLV3* promoter, though the most critical sequences might be in a proximal 350 bp conserved region overlapping the 5' UTR, which was not included in our target region. We also asked if the *slwus*$^{pro}$ alleles could have effects in sensitized genotypes that produce fruits with many locules. To test this, we crossed the *slclv3*$^{pro\text{-}29}$ null allele with two *slwus*$^{pro}$ alleles that disrupted a single CNS or multiple CNSs (*slwus*$^{pro\text{-}4}$ and *slwus*$^{pro\text{-}5}$, respectively), and both double mutants showed a partial suppression of locule number (mean of 12-14 locules in double mutants compared to mean of 16 locules in *slclv3*$^{pro\text{-}29}$) (**Fig. 4h**). These results are also consistent with quantitative epistatic relationships between *cis*-regulatory mutations in *CLV3* and *WUS*[25]. Thus, mutations in the promoter of

*SlWUS* can also cause locule number variation, and some of these effects may depend on background mutations.

In conclusion, we have shown that multiple functional components within a promoter, represented by conserved sequences and their genetic interactions, underlie the complex relationships between *cis*-regulatory regions and their contribution to quantitative phenotypic effects. The additive, redundant and synergistic relationships revealed from our *in vivo* dissection of the *SlCLV3* promoter may reflect a broader principle of how promoters and other *cis*-regulatory regions of plant genes are buffered from genetic perturbations, a theme also gaining support from *in vivo cis*-regulatory dissections in animals[22,35,36]. In plants, the large number of predicted CREs in promoters suggest that such complexity is prevalent, and could involve genetic and physical interactions with 3' regions and also over long distances[8,9]. However, as shown by our mutagenesis of *SlCLV3* and *SlWUS* promoters, *cis*-regulatory complexity can vary substantially between genes, highlighting the need for empirical dissections to understand *cis*-regulatory control and its potential to give rise to quantitative variation. Compared to TFBSs, which are often too abundant and diverse for functional characterizations, CNSs result from purifying selection and often overlap with open chromatin[8], making them prime candidates for *in vivo* functional dissections by genome editing. Indeed, CNSs can reduce the large mutational space of *cis*-regulatory regions and facilitate engineering QTLs by identifying sequences most likely to produce quantitative variation for crop improvement.

Our results also show that the quantitative effects from a particular mutation can be influenced by other mutations, particularly those that are closely linked. *Cis*-regulatory variation is pervasive in related genomes[17–21], and the presence of other linked variants might be affecting characterized QTL mutations, making it challenging to predict the precise quantitative effects of engineered *cis*-regulatory mutations in different genetic backgrounds. Moreover, unlinked mutations in coding or *cis*-regulatory regions can further modify outcomes from *cis*-regulatory engineering[37]. With the deployment of precision genome editing tools[38,39], more dissections of complex interactions between natural and engineered variants that shape quantitative variation will emerge, which in turn will guide the precise design of *cis*-regulatory alleles for crop improvement.

## References


1. Swinnen, G., Goossens, A. & Pauwels, L. Lessons from Domestication: Targeting Cis-Regulatory Elements for Crop Improvement. *Trends in Plant Science* **21**, 506–515 (2016).
2. Meyer, R. S. & Purugganan, M. D. Evolution of crop species: Genetics of domestication and diversification. *Nature Reviews Genetics* **14**, 840–852 (2013).
3. Muños, S. *et al.* Increase in Tomato Locule Number Is Controlled by Two Single-Nucleotide Polymorphisms Located Near WUSCHEL. *Plant Physiol.* **156**, 2244–2254 (2011).
4. Xu, C. *et al.* A cascade of arabinosyltransferases controls shoot meristem size in tomato. *Nat. Genet.* **47**, 784–792 (2015).
5. Somssich, M., Je, B. Il, Simon, R. & Jackson, D. CLAVATA-WUSCHEL signaling in the shoot meristem. *Development* **143**, 3238–3248 (2016).
6. Soyars, C. L., James, S. R. & Nimchuk, Z. L. Ready, aim, shoot: stem cell regulation of the shoot apical meristem. *Curr. Opin. Plant Biol.* **29**, 163–168 (2016).
7. Galli, M., Feng, F. & Gallavotti, A. Mapping Regulatory Determinants in Plants. *Front. Genet.* **11**, 1280 (2020).
8. Lu, Z. *et al.* The prevalence, evolution and chromatin signatures of plant regulatory elements. *Nat. Plants* **5**, 1250–1259 (2019).
9. Ricci, W. A. *et al.* Widespread long-range cis-regulatory elements in the maize genome. *Nat. Plants* **5**, 1237–1249 (2019).
10. Tu, X. *et al.* Reconstructing the maize leaf regulatory network using ChIP-seq data of 104 transcription factors. *Nat. Commun.* **11**, 5089 (2020).
11. Song, B. *et al.* Constrained non-coding sequence provides insights into regulatory elements and loss of gene expression in maize. Preprint at https://doi.org/10.1101/2020.07.11.192575 (2020).
12. Jores, T. *et al.* Identification of Plant Enhancers and Their Constituent Elements by

STARR-seq in Tobacco Leaves. *Plant Cell* **32**, 2120–2131 (2020).
13. Mejía-Guerra, M. K. & Buckler, E. S. A k-mer grammar analysis to uncover maize regulatory architecture. *BMC Plant Biol.* **19**, 103 (2019).
14. Tian, F., Yang, D.-C., Meng, Y.-Q., Jin, J. & Gao, G. PlantRegMap: charting functional regulatory maps in plants. *Nucleic Acids Res.* (2019). doi:10.1093/nar/gkz1020
15. Crisp, P. A. *et al.* Stable unmethylated DNA demarcates expressed genes and their cis-regulatory space in plant genomes. *Proc. Natl. Acad. Sci.* **117**, 23991–24000 (2020).
16. Springer, N., de León, N. & Grotewold, E. Challenges of Translating Gene Regulatory Information into Agronomic Improvements. *Trends Plant Sci.* **24**, 1075–1082 (2019).
17. Alonge, M. *et al.* Major Impacts of Widespread Structural Variation on Gene Expression and Crop Improvement in Tomato. *Cell* **182**, 145-161.e23 (2020).
18. Liu, Y. *et al.* Pan-Genome of Wild and Cultivated Soybeans. *Cell* **182**, 162-176.e13 (2020).
19. Walkowiak, S. *et al.* Multiple wheat genomes reveal global variation in modern breeding. *Nature* **588**, 277–283 (2020).
20. Jayakodi, M. *et al.* The barley pan-genome reveals the hidden legacy of mutation breeding. *Nature* **588**, 284–289 (2020).
21. Haberer, G. *et al.* European maize genomes highlight intraspecies variation in repeat and gene content. *Nat. Genet.* **52**, 950–957 (2020).
22. Fuqua, T. *et al.* Dense and pleiotropic regulatory information in a developmental enhancer. *Nature* **587**, 235–239 (2020).
23. Jakobson, C. M. & Jarosz, D. F. What Has a Century of Quantitative Genetics Taught Us About Nature's Genetic Tool Kit? *Annu. Rev. Genet.* **54**, 439–464 (2020).
24. Zeitlinger, J. Seven myths of how transcription factors read the cis-regulatory code. *Curr. Opin. Syst. Biol.* **23**, 22–31 (2020).
25. Rodríguez-Leal, D., Lemmon, Z. H., Man, J., Bartlett, M. E. & Lippman, Z. B. Engineering Quantitative Trait Variation for Crop Improvement by Genome Editing. *Cell* **171**, 470-480.e8 (2017).
26. Fletcher, J. The CLV-WUS Stem Cell Signaling Pathway: A Roadmap to Crop Yield Optimization. *Plants* **7**, 87 (2018).
27. Fan, C. *et al.* A Novel Single-Nucleotide Mutation in a CLAVATA3 Gene Homolog

Controls a Multilocular Silique Trait in Brassica rapa L. *Mol. Plant* **7**, 1788–1792 (2014).
28. Fletcher, J. C. Signaling of Cell Fate Decisions by CLAVATA3 in Arabidopsis Shoot Meristems. *Science* **283**, 1911–1914 (1999).
29. Suzaki, T. *et al.* Conservation and Diversification of Meristem Maintenance Mechanism in Oryza sativa : Function of the FLORAL ORGAN NUMBER2 Gene. *Plant Cell Physiol.* **47**, 1591–1602 (2006).
30. Rodriguez-Leal, D. *et al.* Evolution of buffering in a genetic circuit controlling plant stem cell proliferation. *Nat. Genet.* **51**, 786–792 (2019).
31. Nelson, A. C. & Wardle, F. C. Conserved non-coding elements and cis regulation: actions speak louder than words. *Development* **140**, 1385–1395 (2013).
32. Laux, T., Mayer, K. F. X., Berger, J. & Jürgens, G. The WUSCHEL gene is required for shoot and floral meristem integrity in Arabidopsis. *Development* **122**, 87–96 (1996).
33. Lin, T.-F., Saiga, S., Abe, M. & Laux, T. OBE3 and WUS Interaction in Shoot Meristem Stem Cell Regulation. *PLoS One* **11**, e0155657 (2016).
34. Mayer, K. F. X. *et al.* Role of WUSCHEL in Regulating Stem Cell Fate in the Arabidopsis Shoot Meristem. *Cell* **95**, 805–815 (1998).
35. Osterwalder, M. *et al.* Enhancer redundancy provides phenotypic robustness in mammalian development. *Nature* **554**, 239–243 (2018).
36. Torbey, P. *et al.* Cooperation, cis-interactions, versatility and evolutionary plasticity of multiple cis-acting elements underlie krox20 hindbrain regulation. *PLOS Genet.* **14**, e1007581 (2018).
37. Soyk, S., Benoit, M. & Lippman, Z. B. New Horizons for Dissecting Epistasis in Crop Quantitative Trait Variation. *Annu. Rev. Genet.* **54**, 287–307 (2020).
38. Ren, Q. *et al.* PAM-less plant genome editing using a CRISPR–SpRY toolbox. *Nat. Plants* **7**, 25–33 (2021).
39. Wang, S. *et al.* Precise, predictable multi-nucleotide deletions in rice and wheat using APOBEC–Cas9. *Nat. Biotechnol.* **38**, 1460–1465 (2020).
40. Brooks, C., Nekrasov, V., Lipppman, Z. B. & Van Eck, J. Efficient gene editing in tomato in the first generation using the clustered regularly interspaced short palindromic repeats/CRISPR-associated9 system. *Plant Physiol.* **166**, 1292–1297 (2014).
41. Naito, Y., Hino, K., Bono, H. & Ui-Tei, K. CRISPRdirect: Software for designing

CRISPR/Cas guide RNA with reduced off-target sites. *Bioinformatics* **31**, 1120–1123 (2015).

42. Werner, S., Engler, C., Weber, E., Gruetzner, R. & Marillonnet, S. Fast track assembly of multigene constructs using golden gate cloning and the MoClo system. *Bioeng. Bugs* **3**, 38–43 (2012).

43. Van Eck, J., Keen, P. & Tjahjadi, M. Agrobacterium tumefaciens-Mediated Transformation of Tomato. in *Transgenic Plants: Methods and Protocols* (eds. Kumar, S., Barone, P. & Smith, M.) 225–234 (Springer New York, 2019). doi:10.1007/978-1-4939-8778-8_16

44. Lemmon, Z. H. *et al.* Rapid improvement of domestication traits in an orphan crop by genome editing. *Nat. Plants* **4**, 766–770 (2018).

45. Qin, C. *et al.* Whole-genome sequencing of cultivated and wild peppers provides insights into Capsicum domestication and specialization. *Proc. Natl. Acad. Sci.* **111**, 5135–5140 (2014).

46. Hardigan, M. A. *et al.* Genome Reduction Uncovers a Large Dispensable Genome and Adaptive Role for Copy Number Variation in Asexually Propagated Solanum tuberosum. *Plant Cell* **28**, 388–405 (2016).

47. Frazer, K. A., Pachter, L., Poliakov, A., Rubin, E. M. & Dubchak, I. VISTA: computational tools for comparative genomics. *Nucleic Acids Res.* **32**, W273–W279 (2004).

48. Mathelier, A. *et al.* JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **44**, D110–D115 (2016).

49. Grant, C. E., Bailey, T. L. & Noble, W. S. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017–1018 (2011).

50. Park, S. J., Jiang, K., Schatz, M. C. & Lippman, Z. B. Rate of meristem maturation determines inflorescence architecture in tomato. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 639–644 (2012).

**Acknowledgements**

We thank members of the Lippman laboratory and D. Jackson for discussions and comments on the manuscript. We thank G. Robitaille, J. Kim, A. Krainer and R. Santos for technical support. We thank Z. H. Lemmon for assistance in allele characterization. We thank L. Randall, A. Doyle, P. Keen, K. Swartwood, M. Tjahjadi and J. Van Eck for performing tomato transformations. We thank T. Mulligan, K. Schlecht, A. Krainer, C. Gilbert, and S. Qiao for assistance with plant care. We thank T. Ha for assistance with statistical analyses. This research was supported by the Howard Hughes Medical Institute and grant support from: the BARD (United States-Israel Binational Agricultural Research and Development Fund, IS-5120-18C) to Z.B.L., and the Vaadia-BARD Postdoctoral Fellowship (award no. FL-542-16) to A.H., and the National Science Foundation Plant Genome Research Program (grant nos. IOS-1546837 and IOS-1732253) to Z.B.L.

**Author contributions**

Z.B.L. conceived the project. X.W., L.A., D.R.L., A.H. and Z.B.L. designed and planned experiments. X.W., L.A., D.R.L., A.H., M.B., and Z.B.L performed experiments and collected data. X.W., L.A., D.R.L., A.H., M.B., and Z.B.L analyzed data, and X.W. and Z.B.L. wrote the manuscript with input from L.A. All authors read, edited, and approved the manuscript.

**Competing interests**

Z.B.L. is a consultant for and a member of the Scientific Strategy Board of Inari Agriculture, and he is also a named inventor on a number of patents and patent applications

directed to related technology that have been exclusively licensed from CSHL to Inari Agriculture.

**Figure legends.**

**Main figures**

**Fig. 1 | A large and diverse collection of CRISPR-Cas9 engineered *SlCLV3* promoter alleles reveals complex relationships between promoter mutations and fruit locule number variation. a**, An expanded collection of 30 *SlCLV3* promoter alleles was generated using a CRISPR-Cas9 genetic drive system[24]. CRISPR-Cas9 transgenic plants are first generated by transforming a construct carrying eight guide RNAs (gRNAs) targeting the promoter. Plants confirmed to carry the Cas9 transgene are then screened by PCR for promoter mutations. Plants biallelic for promoter mutations are then crossed to wild type (WT) plants. The inherited Cas9 transgene can target the WT allele in F1 progeny to generate new mutations. F2 progenies resulting from F1 self-fertilization are then screened by PCR for homozygosity and absence of the Cas9 transgene (i.e. negative for Cas9), and new *SlCLV3* promoter alleles are validated by Sanger sequencing. **b**, Schematics depicting 29 CRISPR-Cas9 engineered *SlCLV3* promoter (*slclv3*$^{pro}$) alleles, along with the domestication QTL allele *fas*. Large deletions, insertions, and inversions are represented by red dashed lines, orange boxes, and red boxes, respectively. Small insertions and deletions (indels) are indicated by numbers and letters. Red arrowheads, gRNAs. Gray line, promoter region. Black box, start of the first exon. **c,** Heatmap representation of the *slclv3*$^{pro}$ alleles and *fas*. The 2.1 kb promoter region is divided into 20 bp windows. Purple color intensity in each window indicates the ratio of sequence changed (i.e. deleted) relative to WT. Red color indicates inversion. **d**, Quantification of fruit locule number. Box plots show the 25$^{th}$, 50$^{th}$ (median), and 75$^{th}$ percentiles for each genotype (left). Number of fruits quantified (n), mean and standard deviation (sd) are shown. **e**, Fruit images showing ranges of locule number

variation for WT and *slclv3*$^{pro}$ alleles. Images of fruits representing locule number ranges (30$^{th}$ and 70$^{th}$ percentiles) for the selected genotypes: WT, *slclv3*$^{pro4}$, *slclv3*$^{pro11}$, *slclv3*$^{pro17}$, *slclv3*$^{pro21}$, *slclv3*$^{pro23}$ and *slclv3*$^{pro28}$ are shown below images. Alleles are ordered according to phenotypic strength (**b-d**).

**Fig. 2 | Mutations in individual conserved sequences of the *SlCLV3* promoter result in weak effects on locule number. a**, Four sets of gRNAs (colored arrowheads) targeting four blocks of conserved *cis*-regulatory sequences in the *SlCLV3* promoter. mVISTA plots of *CLV3* promoter sequence alignments between *Solanum lycopersicum* (tomato) and three other Solanaceae species (potato, *Solanum tuberosum*; pepper, *Capsicum annuum*; groundcherry, *Physalis grisea*) show four conserved regions labeled R1-R4 (colored shading). Blue regions of mVISTA plots indicate >70% sequence similarity over 100 bp windows. Each region was targeted individually by CRISPR-Cas9. Predicted transcription factor binding sites (TFBSs) at relative profile score thresholds of 95% (See Methods) are shown at the bottom (red triangles). **b**, Schematics depicting five *slclv3*$^{pro}$ alleles with targeted mutations in conserved region R1 (*slclv3*$^{pro\text{-}R1}$), their heatmap representations, and quantification of locule numbers. Blue horizontal bars under allele schematics indicate conserved non-coding sequences (CNSs) that are conserved across tomato, potato, pepper and groundcherry. Stacked bar charts show the percentage of total fruits for each locule number. Box plots show the distribution of locule numbers. Number of fruits (n), mean and standard deviation (sd) of locule number are shown. The *slclv3*$^{pro\text{-}R1\text{-}4}$ and *slclv3*$^{pro\text{-}R1\text{-}5}$ alleles showing significant weak effects are outlined with a red box. **c**, Schematic and heatmap representations of three R2 alleles (*slclv3*$^{pro\text{-}R2}$) and quantification of locule numbers. **d**, Schematic and heatmap representations of three R3 alleles (*slclv3*$^{pro\text{-}R3}$) and quantification of locule numbers. **e**, Schematic and heatmap representations of five R4 alleles (*slclv3*$^{pro\text{-}R4}$) and quantification of locule numbers. Box plots in (**b-e**) show the 25$^{th}$, 50$^{th}$, and 75$^{th}$ percentiles for each genotype. The *slclv3*$^{pro\text{-}R4\text{-}4}$ and *slclv3*$^{pro\text{-}R4\text{-}5}$ alleles showing significant weak effects are outlined with a red box. Significantly different locule numbers compared to WT are indicated in (**b**) and (**e**) (p values of two-sided Dunnett's 'compare with control' test less than 0.2 are shown, ns: not significant).

**Fig. 3 | Combining mutations in conserved *cis*-regulatory regions reveals additive, redundant, and synergistic relationships between sequences in the *SlCLV3* promoter. a**, Combinations of mutations in R1 and R4 show increased locule numbers compared to individual mutations in these regions. Schematics and heatmap representations of alleles with mutations in R1 or R4 alone (left-top), and alleles with combined mutations in R1 and R4 from trans-targeting (left-middle) or sequential editing (left-bottom). In alleles with combined mutations, original mutations in one region is labeled by superscript numbers and newly generated mutations in the other region is designated by superscript letters (**Extended Data Fig. 3a,b**). Stacked bar charts and box plots show locule number quantifications (right). **b,** Summary of tests for non-additive effects in combined alleles compared to individual mutations in R1 and R4 regions. If the increase of locule numbers in a combined allele of R1 and R4 is significantly greater than the sum of increases in individual R1 and R4 alleles (adjusted p-values<0.05), then there is a synergistic relationship between the combined mutations. Otherwise, their relationship is additive. Combined allele $R1^{5}$+$R4^{d}$ (labeled with *) showed a non-additive effect in a different experiment (**Extended Data Fig. 3c,d**). **c**, Schematics and heatmap representations of promoter alleles with mutations in R1 or R2 alone and alleles with combined mutations in R1 and R2. Stacked bar charts and box plots show locule number quantifications. **d**, Schematics and heatmap representations of promoter alleles with mutations in R2 or R4 alone and alleles with combined mutations in R2 and R4. Stacked bar charts and box plots show locule number quantifications. **e**, Summary of the genetic relationships between conserved *cis*-regulatory regions. Phenotypic effects of representative alleles with combined mutations showing different genetic relationships (top). A diagrammatic summary of different genetic relationships between conserved *SlCLV3* promoter regions (bottom). Box plots in (**a, c-e**) show the 25$^{th}$, 50$^{th}$, and 75$^{th}$ percentiles for each genotype. P values in (**a, c,** and **d**) are from two-sided Dunnett's 'compare with control' test (p values less than 0.2 are shown, ns: not significant).

**Fig. 4 | The promoter of *SlWUS* is more tolerant to genetic perturbations. a**, CLV3 and WUS function in a conserved negative feedback circuit that modulates stem cell proliferation and meristem size. Red and blue colored areas indicate conventional expression domains for *CLV3* and *WUS*, respectively. LP, leaf primordia. **b**, CRISPR-Cas9 mutagenesis of the *SlWUS* coding sequence ($slwus^{CR\text{-}cds}$). *SlWUS* gene model and gRNA target positions (top, red arrows)

are shown, along with sequences of WT and two *slwus*$^{CR\text{-}cds}$ null alleles (bottom). Black box, black line and grey box represent exon, intron and UTR. **c**, Terminated primary shoot meristem in the *slwus*$^{CR\text{-}cds\text{-}1}$ mutant (right) compared to WT (left). White arrowhead marks terminated meristem with two leaves. L, leaf. The same phenotype was observed for *slwus*$^{CR\text{-}cds\text{-}2}$ (n > 10 individual plants). **d**, Repetitive meristem initiation and termination phenotype of the *slwus*$^{CR\text{-}cds\text{-}1}$ null mutant. Inset in the middle image shows reinitiated disorganized meristems (white arrowheads) that quickly terminate after generating one or two leaf primordia (e.g. red arrowhead). The same phenotype was observed for *slwus*$^{CR\text{-}cds\text{-}2}$. **e**, Schematic depicting seven *SlWUS* promoter (*slwus*$^{pro}$) alleles (top). Blue arrows, gRNA targets. Red dashed lines, deletions. Red box, inversion. Orange triangle, insertion. Predicted TFBSs at relative profile score thresholds of 99% are shown as red triangles (middle). mVISTA plots of *WUS* promoter sequence alignments between tomato and potato, pepper, and groundcherry show five regions of conserved sequences (bottom). **f**, Heatmap representations of *slwus*$^{pro}$ alleles (top). Schematic depicting a CRISPR-Cas9 generated allele of the *SlWUS* 3' region mimicking the domestication *cis*-regulatory QTL allele *lc*[24]. **g**, Quantification of locule number in *slwus*$^{pro}$ alleles. The *slwus*$^{pro\text{-}6}$ allele showing a weak gain-of-function effect is outlined with a red box. **h**, Locule number quantifications showing that *slwus*$^{pro\text{-}4}$ and *slwus*$^{pro\text{-}5}$ reduce locule number in the *slclv3*$^{pro\text{-}29}$ (7.3 kb deletion) background. Alleles are ordered the same in **e-g**. Box plots in **g** and **h** show the 25$^{th}$, 50$^{th}$, and 75$^{th}$ percentiles for each genotype. P values in (**g**) and (**h**) are from two-sided Dunnett's 'compare with control' test (WT and *slclv3*$^{pro\text{-}29}$ as controls, respectively; p values less than 0.2 are shown; ns: not significant).

## Methods

**Plant material, growth conditions, and phenotyping.** Seeds of wild type (*Solanum lycopersicum* cultivar M82, LA3475), *fas*, *slclv3-10* and $lc^{CR}$ in the M82 background were from our own stocks. Seeds were either germinated on moistened filter paper at 28 °C in the dark and later transferred to soil or directly sown in soil in 96-cell plastic flats and grown to 4~5-week-old seedlings in the greenhouse before being transplanted to pots in the greenhouse or directly to fields at Cold Spring Harbor Laboratory. The greenhouse condition is long-day (16 h light, 26-28 °C / 8 h dark, 18-20 °C; 40-60% relative humidity) with natural light supplemented with artificial light from high-pressure sodium bulbs (~250 μmol $m^{-2}$ $s^{-1}$). Plants in the fields were grown under drip irrigation and standard fertilizer regimes, and were used for quantifications of fruit locule number. We counted locules from approximately 100 fruits from about 10 individual plants for each genotype. The locule number phenotyping experiments were repeated over two summer field seasons, representing different soil conditions and environments. Locule data in Fig. 1, Extended Data Fig. 2, Extended Data Fig. 3 and Extended Data Fig. 4 are from experiments in 2019, while data in Fig. 2, Fig. 3 and Fig. 4 and Extended Data Fig. 1 are from experiments in 2020. Phenotypes of *slwus* null mutants were observed in more than ten plants during at least two growing seasons.

**CRISPR–Cas9 mutagenesis, plant transformation, and selection of mutant alleles.** CRISPR-Cas9 mutagenesis and generation of transgenic tomato lines were performed as described previously [40]. Briefly, gRNAs were designed using the CRISPRdirect tool (https://crispr.dbcls.jp/)[41]. Binary vectors for Cas9 and gRNAs were assembled using Golden Gate cloning as described [25,42]. The final binary plasmids were introduced into wild type M82 or homozygous promoter alleles by *Agrobacterium tumefaciens*-mediated transformation through tissue culture [43]. First-generation (T0) transgenic plants were transplanted in soil and grown under standard greenhouse conditions. Genotyping of CRISPR-generated mutations was performed as previously described [25]. Briefly, gRNA target regions were PCR amplified in T0 transgenic plants (gRNA and primer sequences for genotyping are listed in Supplementary Table 1). PCR products were then analyzed by gel electrophoresis and cloned into pSC-B-amp/kan (Agilent) following the manufacturer's instructions for Sanger sequencing. Sequences were assembled using Geneious (v11.1.5).

**Cis-regulatory sequence conservation analyses and TFBS prediction.** For comparative sequence analysis of Solanaceae *CLV3* promoters, the syntenic regions of *SlCLV3* and surrounding sequences in *S. tuberosum*, *S. annuum* and *P. grisea* were identified by BLAST using the *SlCLV3* genomic sequence, including the protein coding regions[44–46]. 3 kb genomic sequences upstream of the *CLV3* coding regions from *S. tuberosum*, *S. annum* and *P. grisea* were aligned to those of *S. lycopersicum* using mVISTA LAGAN alignment (http://genome.lbl.gov/vista/mvista/submit.shtml )[47]. The plots show alignment windows of 100bp at a similarity threshold of 70%, highlighted in blue. The same analysis was performed with 3kb promoter sequences of *WUS*. Predicted TFBSs were identified from 1.5 kb of the *SlCLV3* promoter and 2.6 kb of the *SlWUS* promoter. Plant TF motifs in JASPAR Core Plantae[48] were used with FIMO motif scanning in the MEME suite (http://meme-suite.org/doc/fimo.html)[49]. Relative profile score thresholds of 95% and 99% were used as cut-offs to show TFBSs in the *SlCLV3* and *SlWUS* promoters, respectively.

**RNA extraction and Quantitative RT-PCR (qPCR).** For gene expression analysis, seeds were germinated on moistened filter paper at 28 °C in dark. After germination, seedlings at similar stages were transferred to soil in 96-cell plastic flats and grown in the greenhouse. Shoot apices including the first floral meristem and sympodial inflorescence meristems were collected at the floral meristem stage of meristem maturation[50], and immediately flash-frozen in liquid nitrogen. Seven to ten apices were combined as one biological replicate and three replicates were collected for each genotype. Total RNA was extracted using TRIzol® Reagent (Invitrogen) and 200 ng of total RNA was used for cDNA synthesis using the SuperScript IV VILO Master Mix (Invitrogen). qPCR was performed with gene-specific primers using the iQ SYBR Green SuperMix (Bio-Rad) reaction system on the CFX96 Real-Time system (Bio-Rad). Primer sequences are available in Supplementary Table 1.

**Statistical analyses.** For Pairwise comparisons between promoter alleles and wild type, locule number phenotypes in alleles having mutations in individual conserved regions or with combined mutations were compared to the isogenic wild type control M82 using Dunnett's ‘compare with control’ tests.

For tests of genetic interactions, a pairwise interaction between mutations in two regions (e.g. R1 and R4) was defined as the difference between the locule number change in the combined allele, *R1+R4*, and the expected locule number change obtained by the addition of locule number changes from alleles with mutations in single regions (R1 and R4). The interaction between two mutations is:

$$\begin{aligned}\varepsilon_{R1+R4} &= (\mu_{R1+R4} - \mu_{WT}) - \big((\mu_{R1} - \mu_{WT}) + (\mu_{R4} - \mu_{WT})\big) \\ &= \mu_{R1+R4} - \mu_{R1} - \mu_{R4} + \mu_{WT}\end{aligned}$$

in which $\mu$ is the mean locule number. To test if there was any significant interaction, the probability of $\varepsilon_{R1+R4}$ being different from 0 (the p-value) was calculated using the parameters below:

The sample distribution of $\varepsilon_{R1+R4}$ follows approximately a normal distribution with mean estimate

$$\hat{\mu} = \hat{\varepsilon}_{R1+R4} = \hat{\mu}_{R1+R4} - \hat{\mu}_{R1} - \hat{\mu}_{R4} + \hat{\mu}_{WT}$$

and variance of

$$\hat{\sigma}^2 = \frac{\hat{\sigma}_{R1+R4}{}^2}{n_{R1+R4}} + \frac{\hat{\sigma}_{R1}{}^2}{n_{R1}} + \frac{\hat{\sigma}_{R4}{}^2}{n_{R4}} + \frac{\hat{\sigma}_{WT}{}^2}{n_{WT}}$$

in which $\hat{\sigma}$ is the sample variance and n is number of samples. P values were adjusted using Benjamini-Hochberg (BH) method. Since newly induced mutations in the combined alleles are not exactly the same as those in alleles with mutations only in individual conserved regions, we used the strongest phenotypes from mutations in each CNS region to represent their expected additive effects and tested for non-additivity.

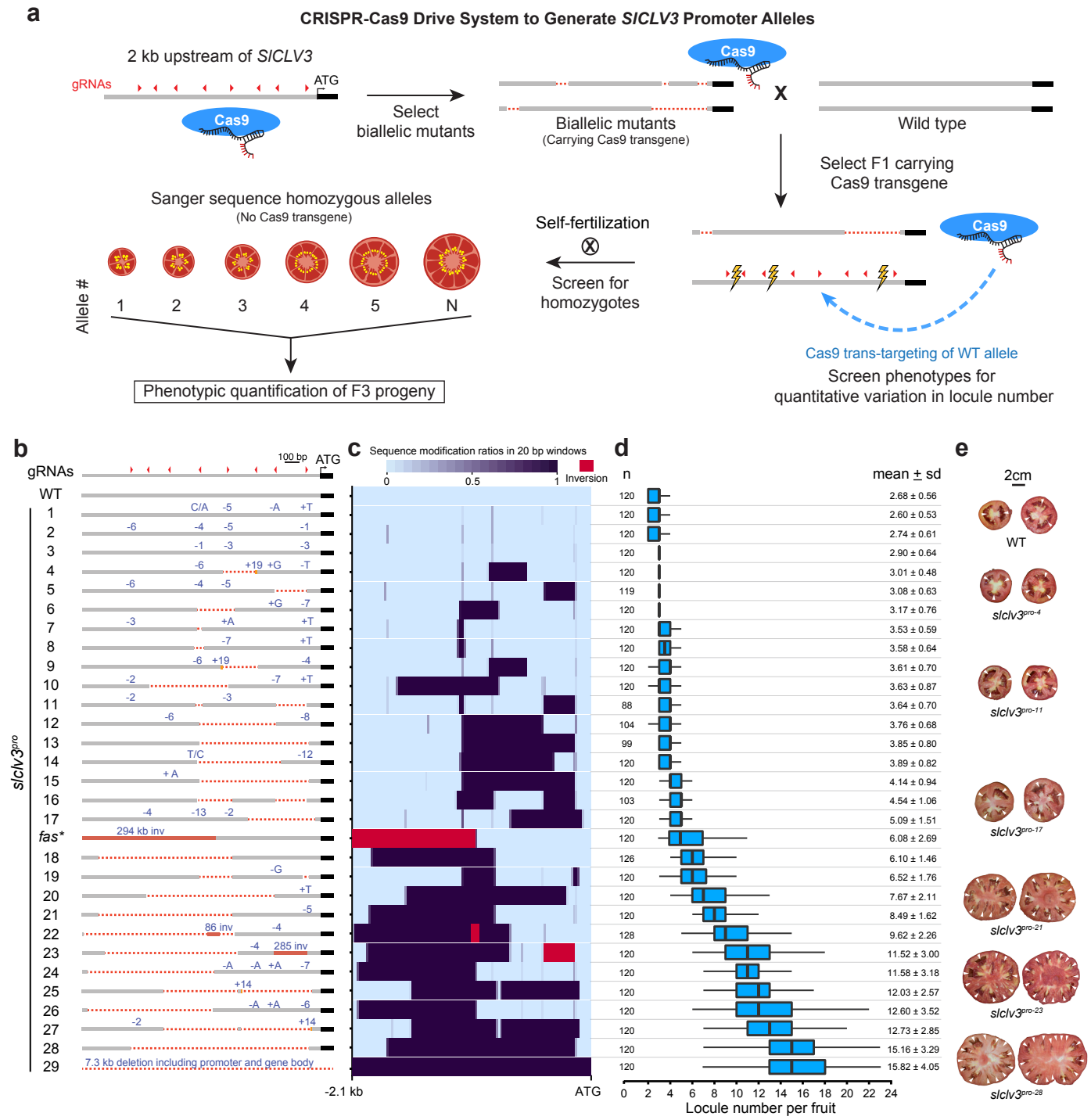
For expression analyses using RT-qPCR, three biological replicates of pooled meristems were used for each genotype and at least two technical replicates were performed for each biological replicate. Means ± s.e. were shown and mean values between groups were compared by two-sample t tests.

**Reporting Summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

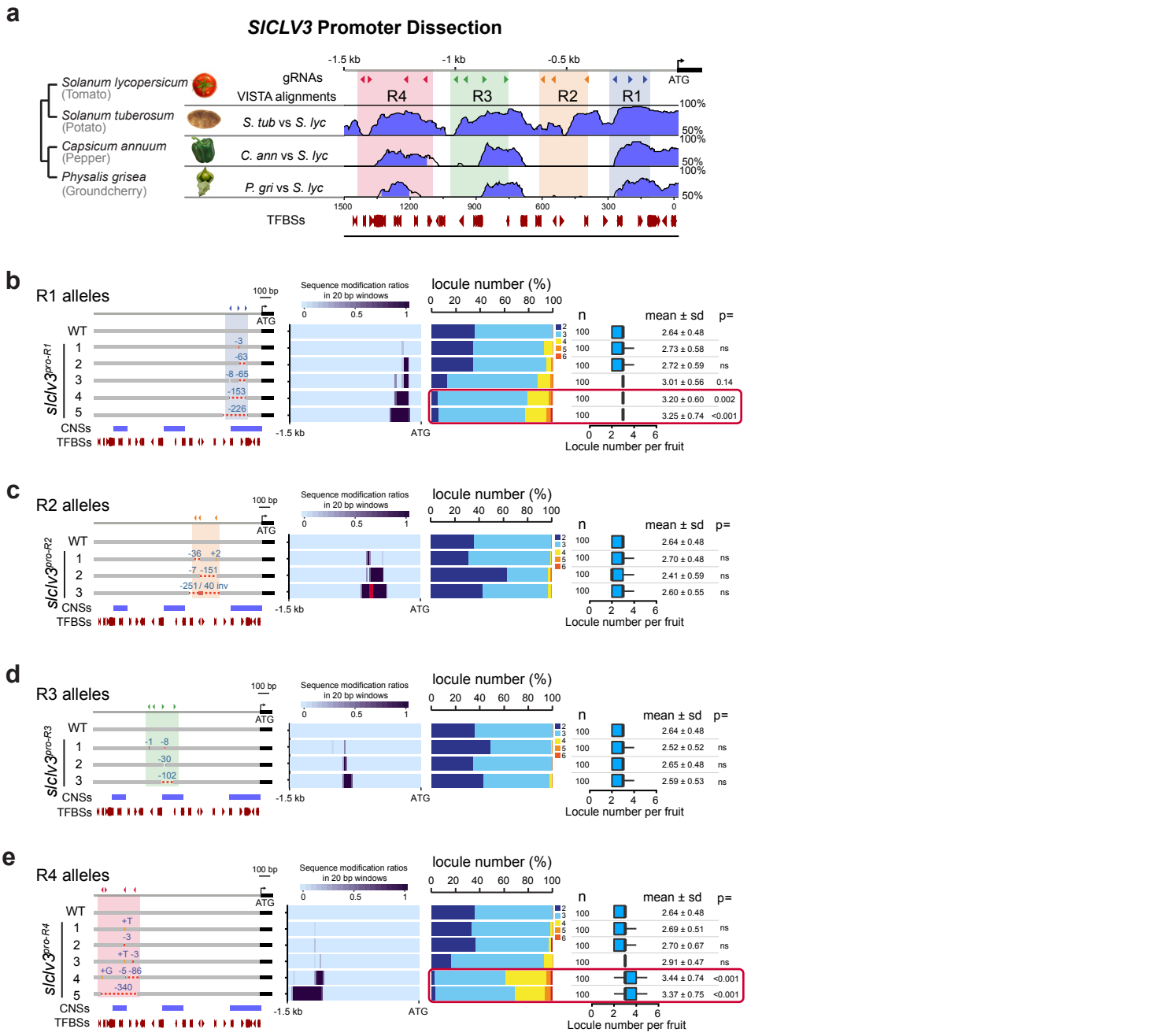
## Data Availability

Source Data files for all main and Extended Data figures are available in the online version of the paper. All additional data sets are available from the corresponding author upon request.

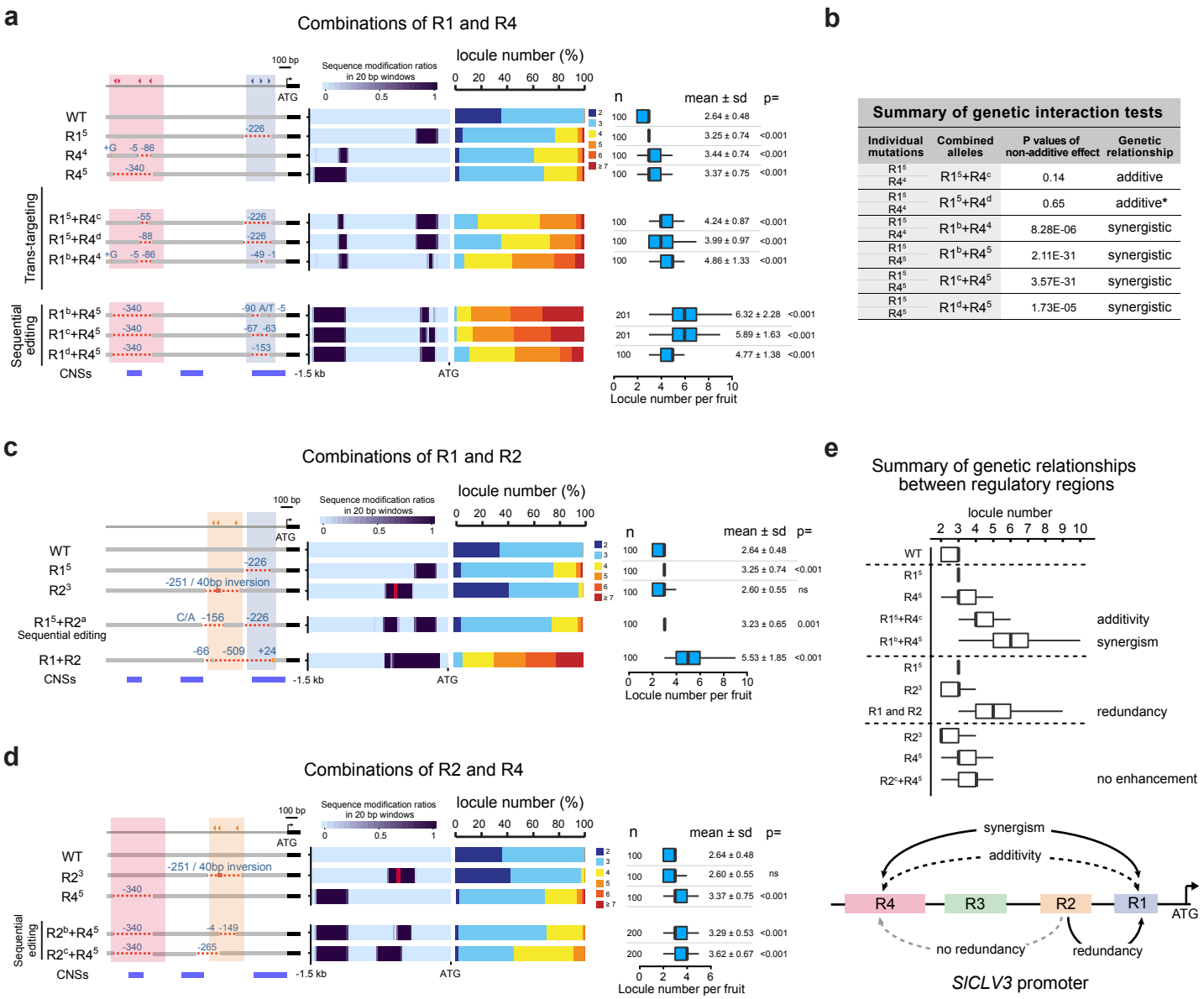
# Figure 1


a
CRISPR-Cas9 Drive System to Generate SICLV3 Promoter Alleles
2 kb upstream of SlCLV3
gRNAs
ATG
Cas9
Select biallelic mutants
Biallelic mutants
(Carrying Cas9 transgene)
X
Wild type
Select F1 carrying Cas9 transgene
Sanger sequence homozygous alleles
(No Cas9 transgene)
Allele #
1
2
3
4
5
N
Self-fertilization
Screen for homozygotes
Cas9 trans-targeting of WT allele
Screen phenotypes for quantitative variation in locule number
Phenotypic quantification of F3 progeny
b
gRNAs
WT
slclv3pro
fas*
294 kb inv
7.3 kb deletion including promoter and gene body
86 inv
285 inv
c
Sequence modification ratios in 20 bp windows
Inversion
-2.1 kb
ATG
d
n
mean ± sd
Locule number per fruit
e
2cm
WT
slclv3pro-4
slclv3pro-11
slclv3pro-17
slclv3pro-21
slclv3pro-23
slclv3pro-28

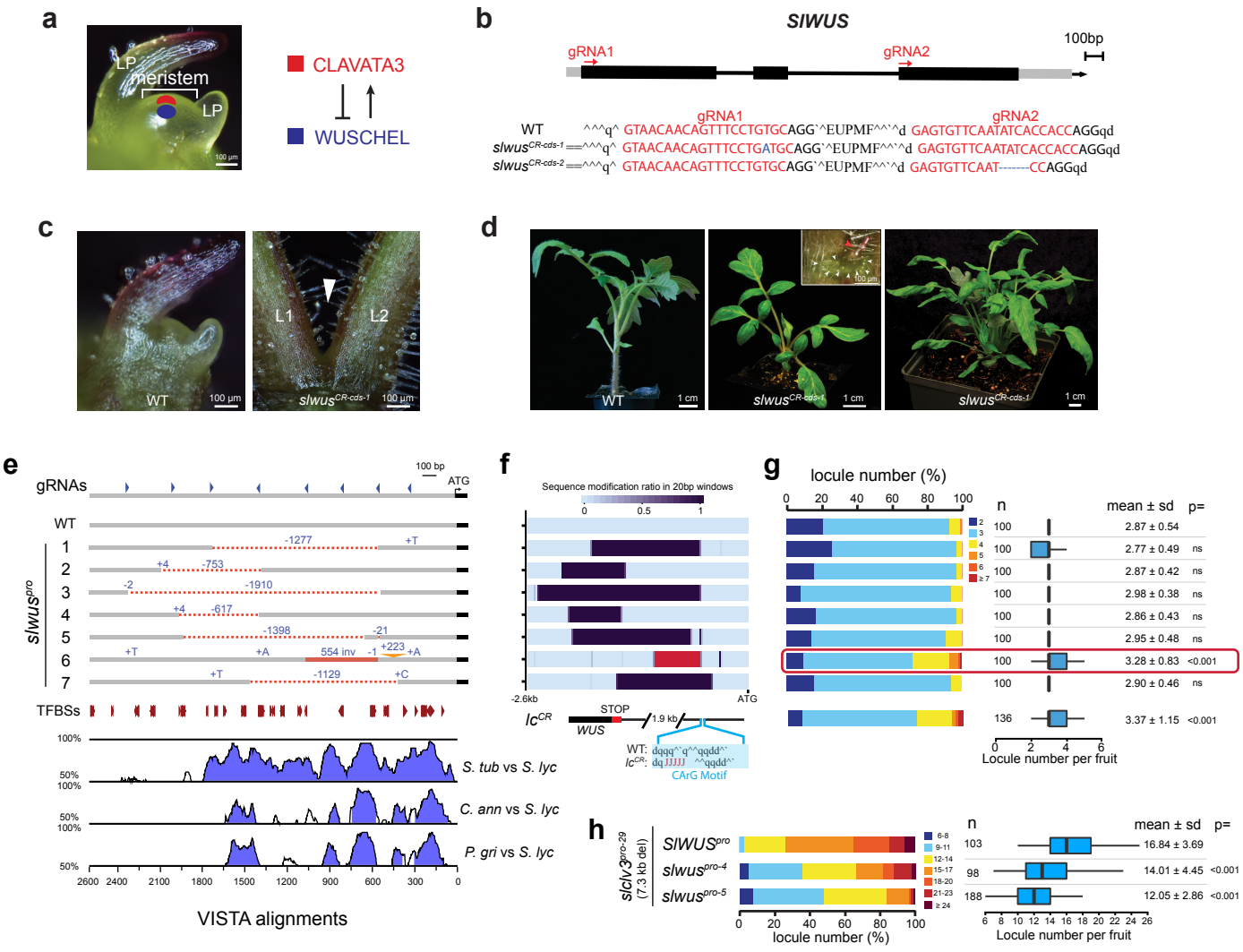
# Figure 2


a
SlCLV3 Promoter Dissection
Solanum lycopersicum (Tomato)
Solanum tuberosum (Potato)
Capsicum annuum (Pepper)
Physalis grisea (Groundcherry)
gRNAs
VISTA alignments
S. tub vs S. lyc
C. ann vs S. lyc
P. gri vs S. lyc
TFBSs
R4
R3
R2
R1
-1.5 kb
-1 kb
-0.5 kb
ATG
b
R1 alleles
c
R2 alleles
d
R3 alleles
e
R4 alleles
Sequence modification ratios in 20 bp windows
locule number (%)
n
mean ± sd
p=
Locule number per fruit
CNSs
TFBSs

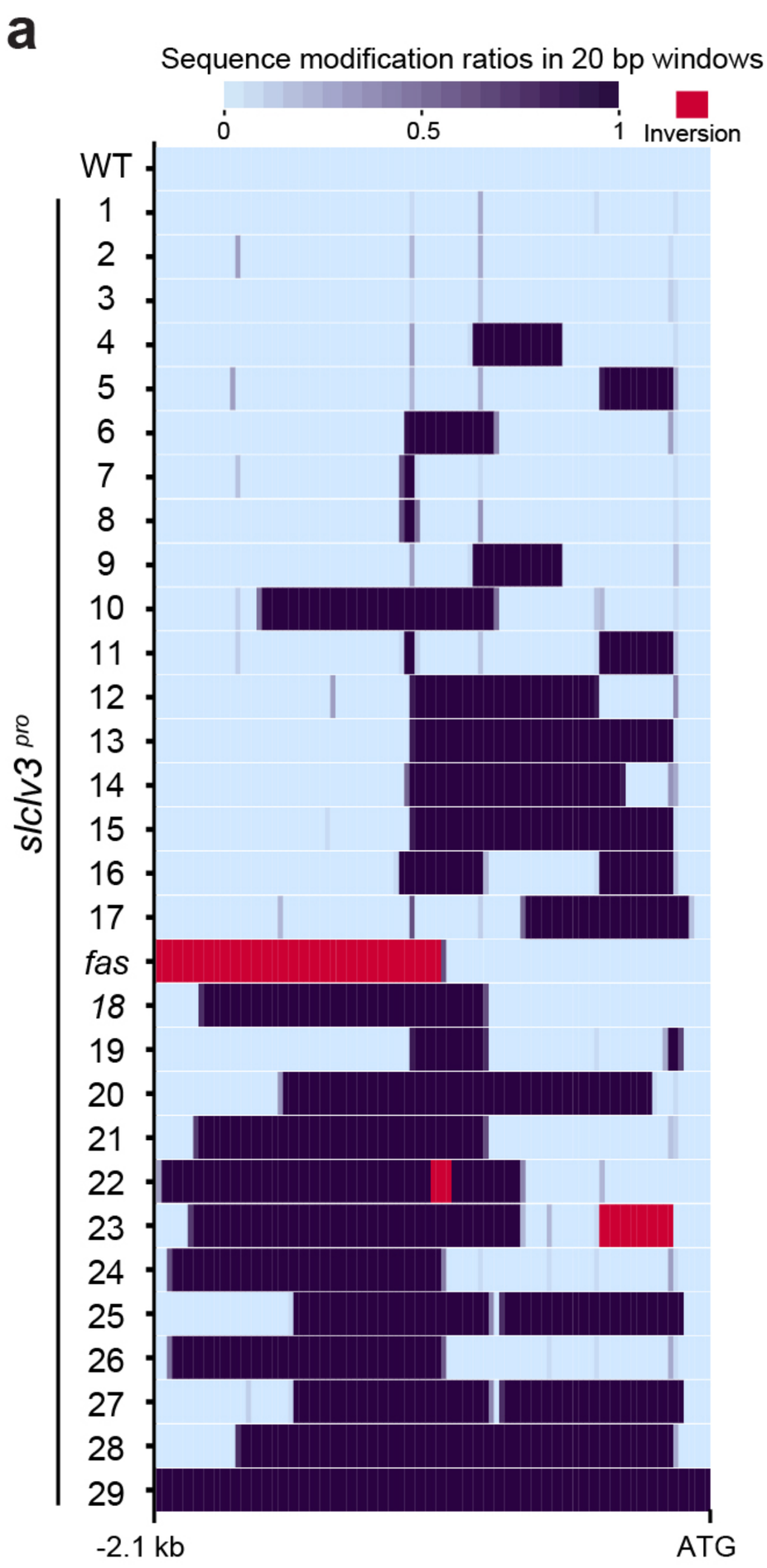
# Figure 3

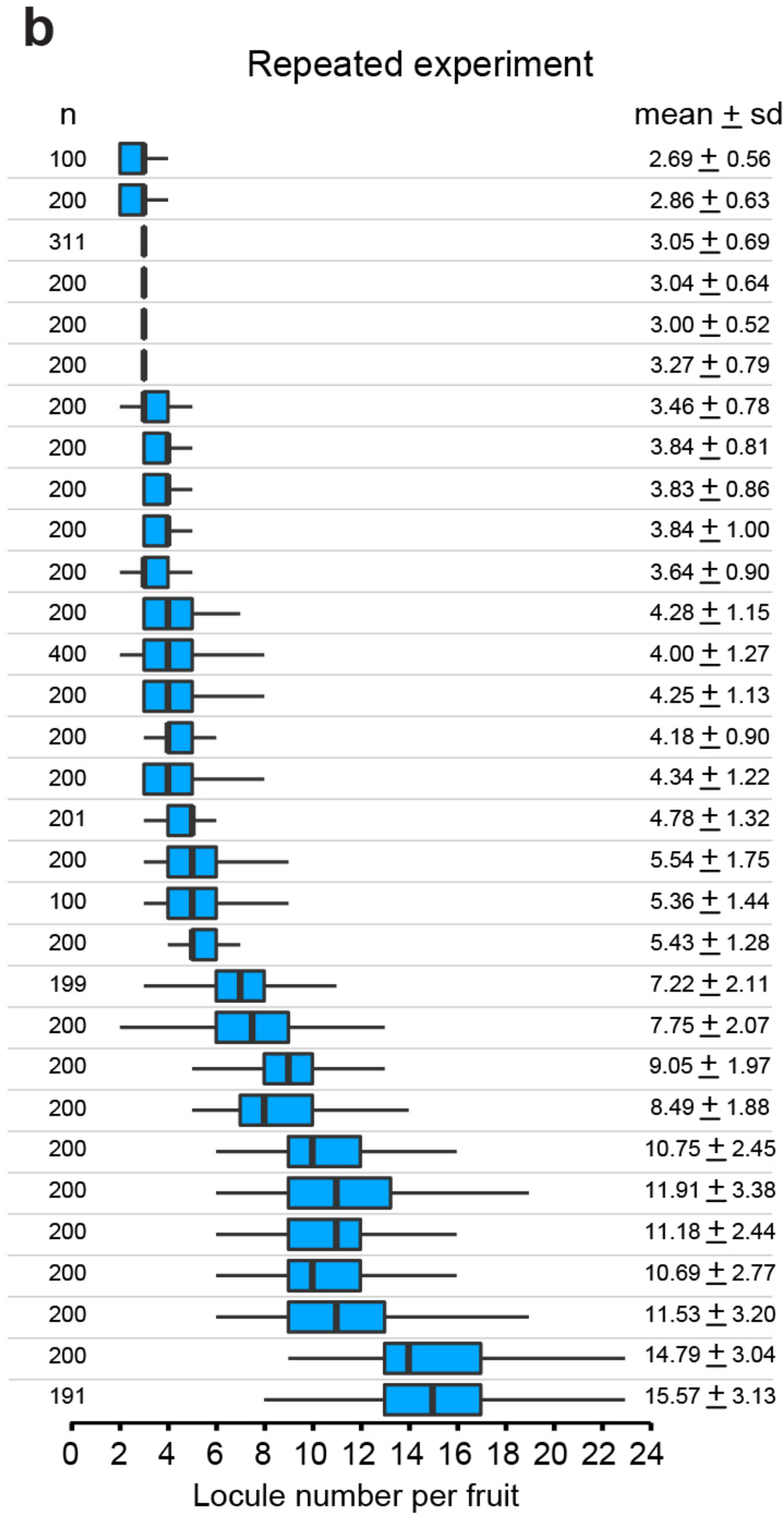

a
Combinations of R1 and R4
100 bp
ATG
Sequence modification ratios in 20 bp windows
locule number (%)
n
mean ± sd
p=
WT
R1^5
R4^4
R4^5
Trans-targeting
R1^5+R4^c
R1^5+R4^d
R1^b+R4^4
Sequential editing
R1^b+R4^5
R1^c+R4^5
R1^d+R4^5
CNSs
-1.5 kb
ATG
Locule number per fruit
b
Summary of genetic interaction tests
Individual mutations
Combined alleles
P values of non-additive effect
Genetic relationship
R1^5 R4^4 R1^5+R4^c 0.14 additive
R1^5 R4^4 R1^5+R4^d 0.65 additive*
R1^5 R4^4 R1^b+R4^4 8.28E-06 synergistic
R1^5 R4^5 R1^b+R4^5 2.11E-31 synergistic
R1^5 R4^5 R1^c+R4^5 3.57E-31 synergistic
R1^5 R4^5 R1^d+R4^5 1.73E-05 synergistic
c
Combinations of R1 and R2
R2^3
R1^5+R2^a
Sequential editing
R1+R2
d
Combinations of R2 and R4
R2^b+R4^5
R2^c+R4^5
e
Summary of genetic relationships between regulatory regions
locule number
additivity
synergism
redundancy
no enhancement
R1 and R2
R4
R3
R2
R1
no redundancy
SICLV3 promoter

# Figure 4

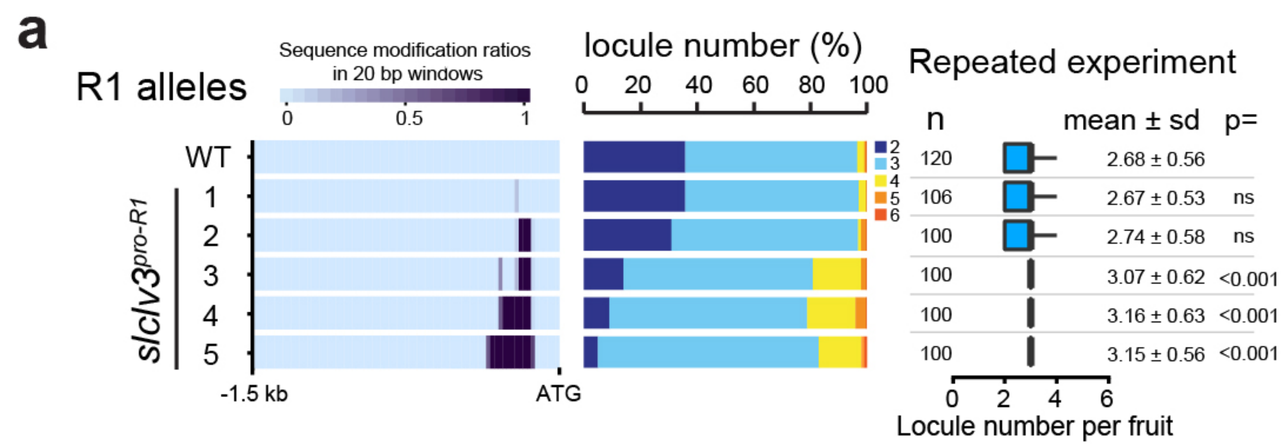

a
LP
meristem
LP
CLAVATA3
WUSCHEL
b
SlWUS
gRNA1
gRNA2
100bp
WT
slwus CR-cds-1
slwus CR-cds-2
c
WT
L1
L2
slwus CR-cds-1
d
WT
slwus CR-cds-1
slwus CR-cds-1
e
gRNAs
ATG
WT
slwus pro
TFBSs
S. tub vs S. lyc
C. ann vs S. lyc
P. gri vs S. lyc
VISTA alignments
f
Sequence modification ratio in 20bp windows
lc CR
STOP
WUS
CArG Motif
g
locule number (%)
n
mean ± sd
p=
Locule number per fruit
h
slclv3 pro-29
SlWUS pro
slwus pro-4
slwus pro-5
locule number (%)
Locule number per fruit

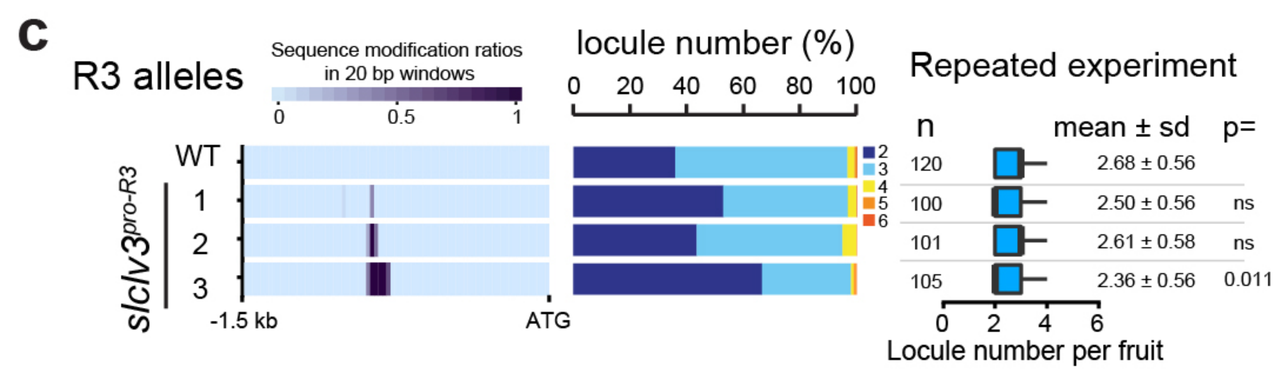
# Extended Data Fig. 1

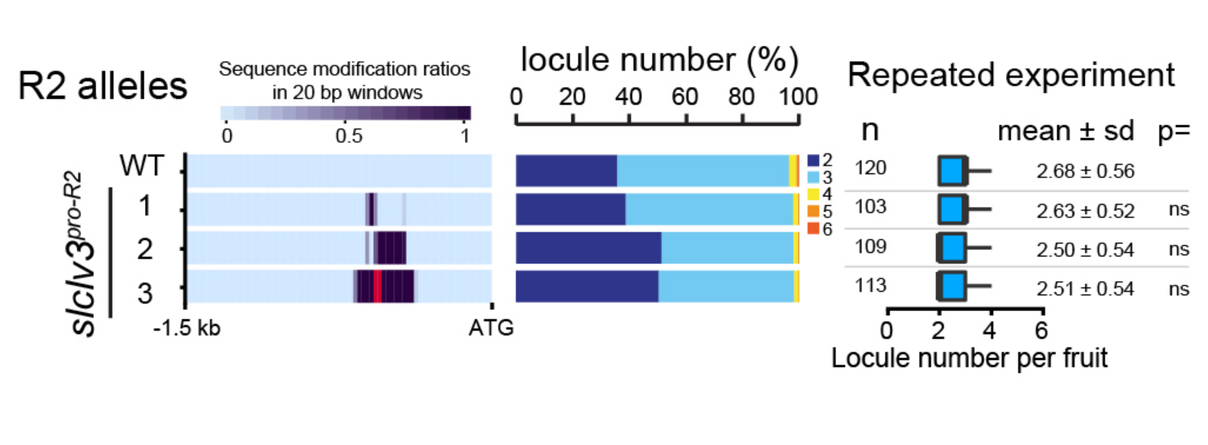


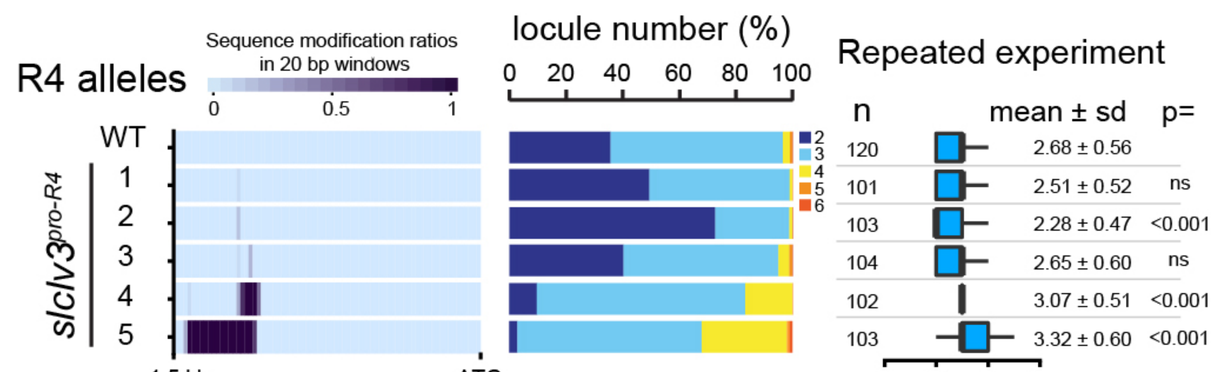



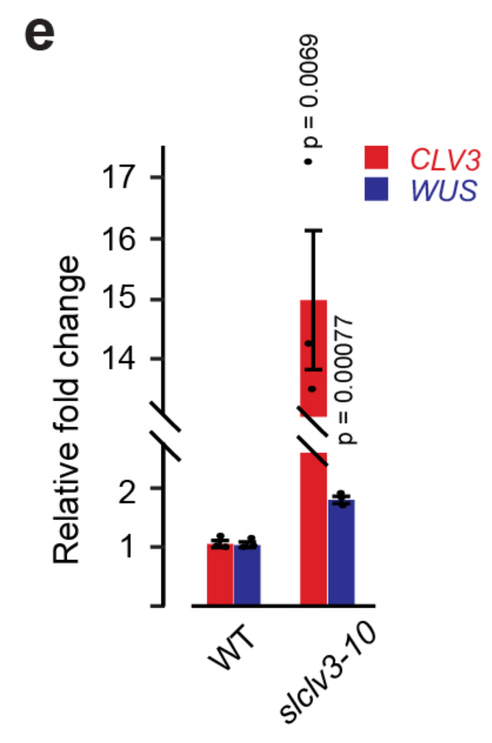
# Extended Data Fig. 2

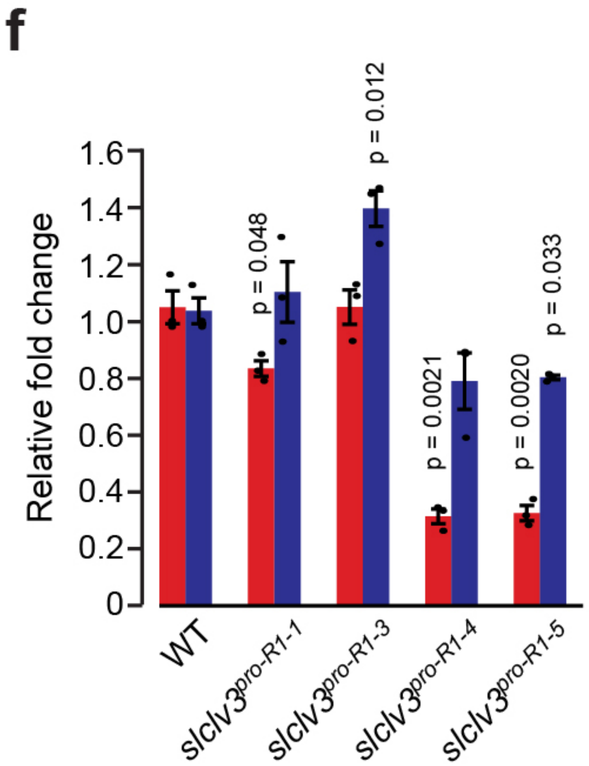


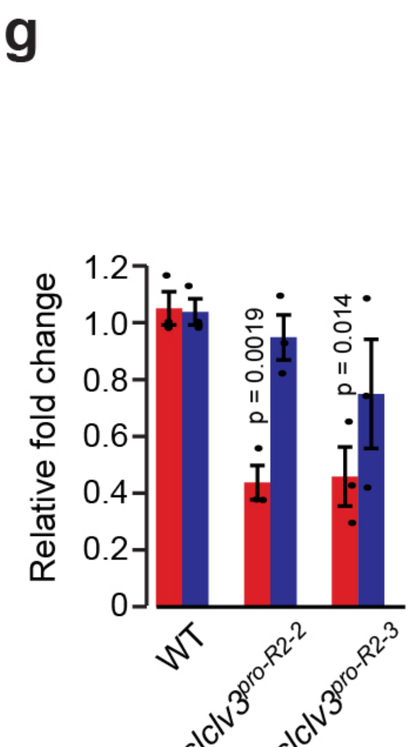



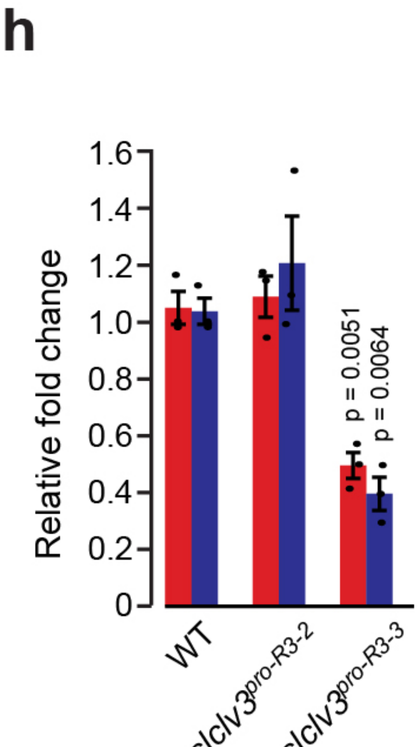



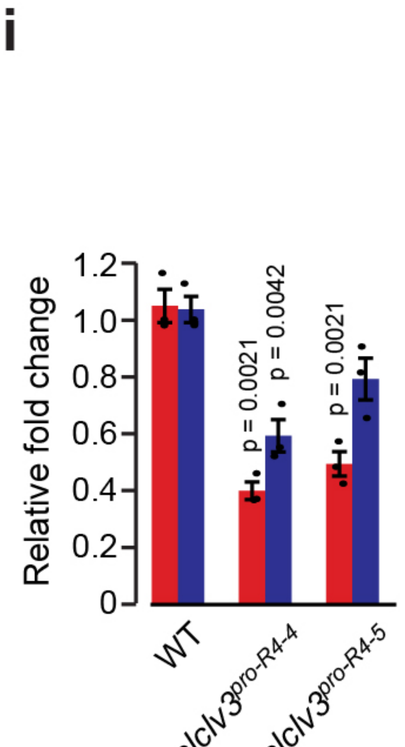



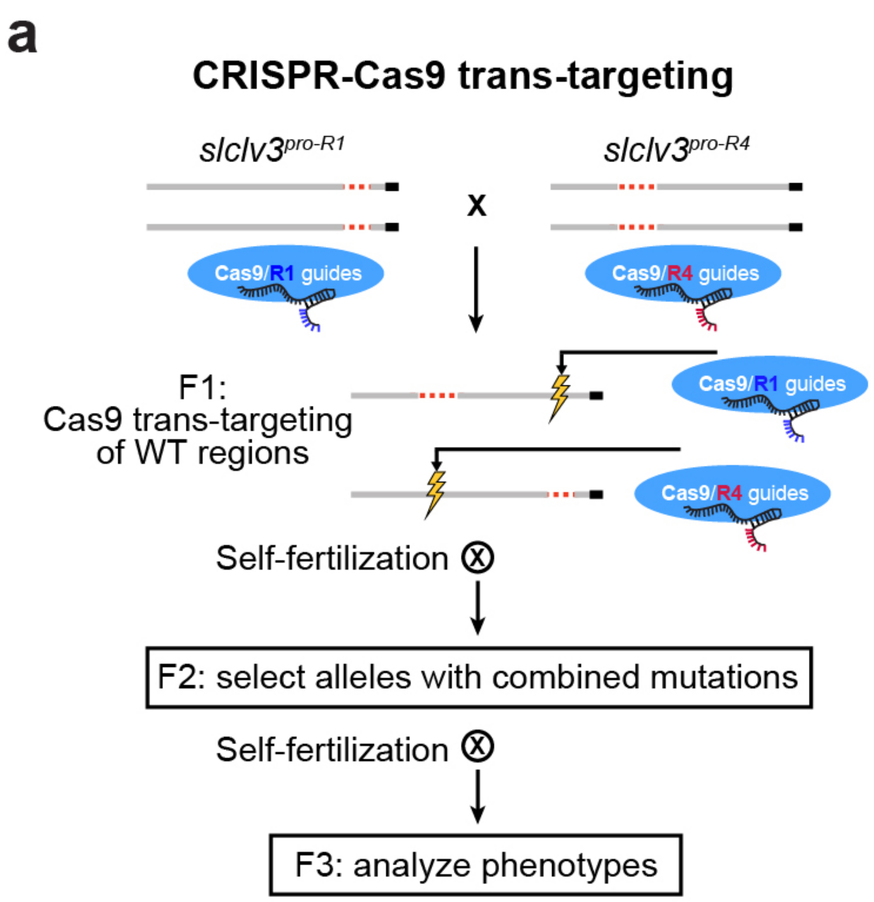



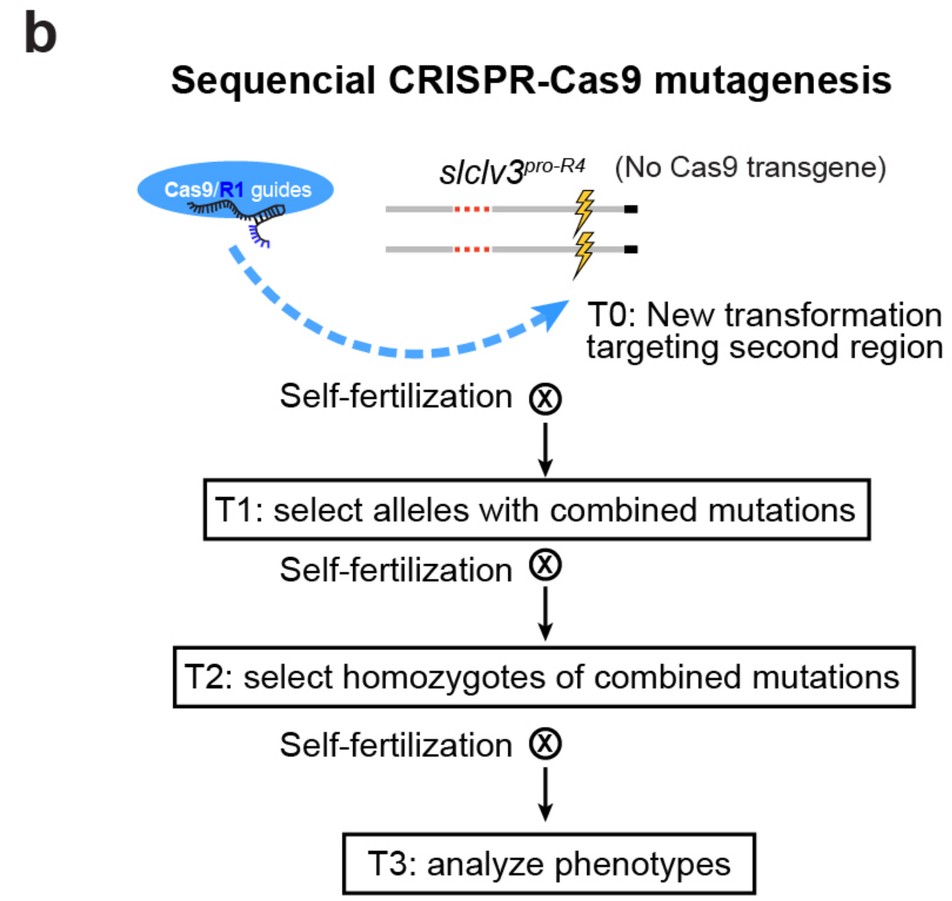



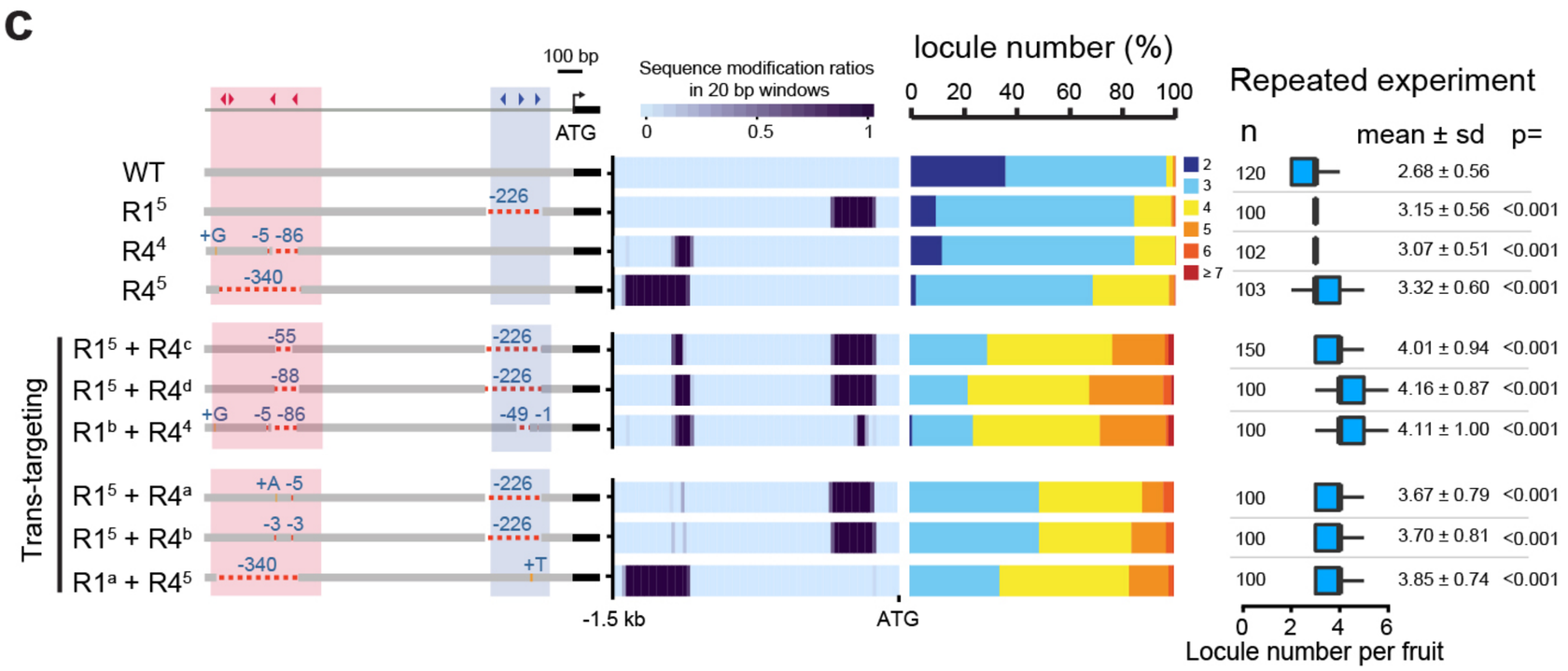



# Extended Data Fig. 3



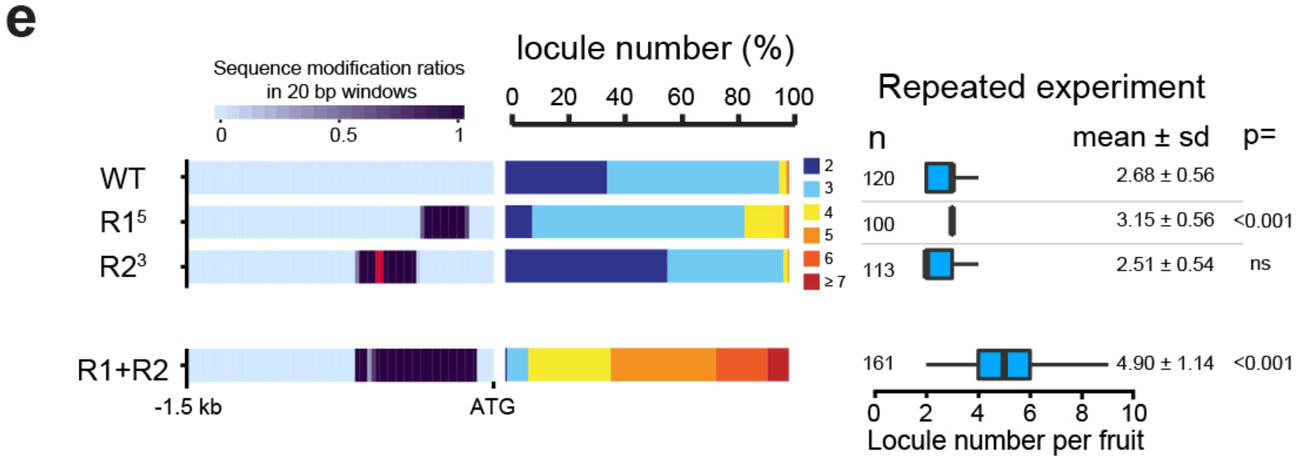




d

**Summary of genetic interaction tests (2019)**

| Individual mutations | Combined alleles | P values of non-additive effect | Genetic relationship |
|---|---|---|---|
| $R1^5$ $R4^5$ | $R1^5$+$R4^c$ | 0.0630 | additive |
| $R1^5$ $R4^5$ | $R1^5$+$R4^d$ | 0.0053 | synergistic* |
| $R1^5$ $R4^5$ | $R1^b$+$R4^4$ | 0.0209 | synergistic |
| | | | |
| $R1^5$ $R4^5$ | $R1^5$+$R4^a$ | 0.2247 | additive |
| $R1^5$ $R4^5$ | $R1^5$+$R4^b$ | 0.2829 | additive |
| $R1^5$ $R4^5$ | $R1^a$+$R4^5$ | 0.3404 | additive |



# Extended Data Fig. 4