Mutator transposons in *Zea mays* impact transcriptional regulatory networks and underlying gene expression

A DISSERTATION
SUBMITTED TO THE FACULTY OF
THE UNIVERSITY OF MINNESOTA
BY

Erika L. Magnusson

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Dr. Nathan M. Springer

January 2023

ACKNOWLEDGEMENTS

I am very grateful for the opportunity I had to pursue my doctoral degree under the mentorship of Nathan Springer. Nathan, thank you for your commitment to helping me become the scientist I am today– for the continual support and direction you provided. Thank you also for fostering such a great lab culture, as I was fortunate to be surrounded by wonderful people while I was learning. I would like to thank all the past and present Springer Lab members, for the science chats, moral support, and shared knowledge. I would like to express my gratitude to Pete Hermanson for all of the hours he spent helping me generate all of the data included in this thesis– for your availability and willingness to serve everyone in the lab.

I would also like to thank my committee, Candy Hirsch, Katie Greenham, and Clay Carter, who generously provided knowledge and expertise. A huge thanks to Candy Hirsch, the chair of my committee, for her mentorship both personally and professionally throughout my PhD. I would also like to thank the members of the Candy and Corey Hirsch labs, everyone in my Plant and Microbiology Biology (PMB) cohort, and the PMB community as a whole– for all of the shared experiences and friendships made. Additionally, my PhD research would not have been possible without the generous financial support from the National Science Foundation Plant Genome Research Program (PGRP) and collaborations with the Erich Grotewold, Natalia de Leon, Andrea Doseff, and John Gray groups.

I would like to thank my family, especially my parents and sisters, who have supported me throughout my life and educational goals. A special thanks to my late grandfather, Alan “Buddy” Magnusson, and my father Mark for introducing me to agriculture. Dad, thanks for all the agronomy lessons along the way. Lastly, I am most grateful for the support of my husband Evan. We were fortunate to journey through the PMB graduate program together and your belief in me as I am now and who I will become means the most.

THESIS ABSTRACT

Eukaryotic gene expression is transcriptionally regulated by *cis*- and *trans*-regulatory inputs. The interactions between *cis*-regulatory elements (CREs) and transcription factor (TF) *trans*-regulators that recognize CREs coordinate regulation of gene expression. Much work has been done to identify these protein-DNA interactions because they represent interactions that can directly change transcriptional activity. Interactions between TFs and target genes can be predicted by representative gene-regulatory networks (GRNs) to derive these causal relationships. In crop species, the rewiring of GRNs by either selecting for existing variants or introducing new genetic variants is a potential strategy for trait improvement. However, the predictive power of inferred GRNs must be determined experimentally. One method to test whether GRN predicted TF-target gene interactions represent functionally important interactions is to perturb the networks.

In maize, a family of DNA transposons, *Mutator*, can be used to disrupt gene function and are known to contain *cis*-regulatory sequences that may influence gene expression. Transposable elements (TEs) or transposons are mobile repetitive DNA sequences that have proliferated in the genomes of many crop species. To maintain gene function, host genomes have found ways to silence TEs and TEs have adapted to persist regardless of this silencing. TEs may carry potential cryptic regulatory sequences that were once used in their ancestral state to facilitate their own transposition. It is important to determine how these TE-derived CREs have contributed to TE evolution and gene regulation via *cis*-*trans* interactions.

In this thesis we use *Mutator* in maize to study TF-target gene interactions and transposon biology. The central questions we ask in these thesis chapters are (1) what is the relative accuracy of predicted GRNs in maize (2) can we determine the value of GRN predictions by perturbation with TF loss-of-function mutants in maize (3) what is the frequency and function of putative cryptic promoters in *Mutator* (4) can we study *Mutator* to learn more about how the effects of TE insertion can be masked when regulatory sequences within TEs are co-opted by the host genome.

To address these thesis questions, I first present background information on both *Mutator* transposons and their use for reverse genetics studies in maize (Chapter I). It is pertinent to understand why *Mutator* is used in maize for insertional mutagenesis and how properties of *Mutator* can impact downstream analyses. In addition, I will provide brief background information on why it is necessary to determine the functional relevance of inferred GRNs *in vivo*. I will then present two research studies that were conducted in maize for this thesis. TF mutant alleles were isolated from a *Mutator* transposon-indexed population in maize, UniformMu. We utilized these mutant alleles to study how *Mutator* may provide novel alternative promoters that can impact gene expression patterns and levels and influence TE-host genome evolution (Chapter II). We also used these mutant alleles to perturb predicted GRNs in maize and test expression of predicted targets (Chapter III). To summarize these studies, we provide insight into what should be known when using *Mutator* insertion stocks for future reverse genetics studies in maize.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER II

CHAPTER III

LIST OF FIGURES

CHAPTER II

CHAPTER III

LIST OF ABBREVIATIONS

CRE	*Cis*-regulatory element
TF	Transcription factor
GRN	Gene regulatory network
TE	Transposable element
Mu	*Mutator*
TIR	Terminal inverted repeat
ORF	Open reading frame
UTR	Untranslated region
RNA-seq	RNA sequencing
ChIP-seq	Chromatin immunoprecipitation followed by sequencing
DAP-seq	DNA-affinity purification followed by sequencing
Y1H	Yeast-One-Hybrid
CPM	Counts per million
FPKM	Fragments per kilobase of exon per million mapped reads
GCN	Gene co-expression network
DE	Differentially expressed
DEG	Differentially expressed genes
GO	Gene ontology

CHAPTER I

Transcriptional regulation of gene expression in maize: Contributions from *Mutator* transposons and transcription factors

Introduction

Eukaryotic genomes are composed of mostly non-coding DNA. Embedded within this non-coding DNA are *cis*-regulatory elements (CREs) that are recognized by *trans*-regulators, such as transcription factors (TFs), to regulate gene expression [1]. However, for most genes the full complement of *cis*-regulatory sequences and *trans*-factors that influence gene expression have not been characterized. Along with CREs, transposable elements (TEs) make up a large proportion of non-coding DNA [2]. These TEs, as part of the non-coding landscape of the genome, may contain CREs within their sequence [3–7]. For trait improvement efforts in crop species, it has become increasingly important to identify interactions between CREs and TFs that can directly influence regulation of gene expression. For this brief review, we will focus on the relevance of using *Mutator* transposon insertion stocks in maize to understand more about how TE-derived CREs and TF-target interactions influence regulation of gene expression.

TEs are relatively large repetitive genetic elements (ranging from 100-10,000 bp) that can move and replicate within the host genome [8]. Eukaryotic TEs can be divided into two classes based on their transposition intermediates and mobilization mechanism. Class I retrotransposons replicate via an RNA intermediate and integrate into the genome by a "copy and paste" mechanism. Class II elements or DNA transposons replicate and mobilize via a DNA intermediate and "cut and paste" mechanism. TEs are further classified into two categories based on their ability to mobilize on their own, autonomously, or not. Transposons that encode the proteins necessary for transposition are referred to as autonomous, while those that require an autonomous element to transpose are non-autonomous [8].

Transposons have been used as a source of insertional mutagenesis in many plant species, including *Arabidopsis* [9,10], *Zea mays* (maize) [11], and *Medicago* [12]. Most of the transposons used for insertional mutagenesis are Class II elements or DNA transposons. One family of DNA transposons that is highly mutagenic and active exclusively in maize is *Mutator (Mu)* [13–15]. The properties of *Mu* that make this element

so mutagenic include preferential insertion into low copy sequences or genic regions, high transposition frequency, and high forward mutation rate in the germline [16–21]. These properties of *Mu* also make it an excellent system to disrupt genes for reverse-genetics studies. Several populations of *Mu*-indexed stocks have been developed in maize and are publicly available [22–25]. In maize, *Mu*-insertion stocks are used as a main resource for isolating loss-of-function mutations in many genes.

The genetic properties of *Mu* that have been advantageous for insertional mutagenesis can also be utilized to study DNA transposon-genome dynamics. Due to the preference of *Mu* to insert into or near genes, like some other DNA transposon families, *Mu* can provide a system to study how transposons contribute to the evolution of gene regulation. Here we will describe more about what is known of *Mu* behavior and regulation, mechanisms for how *Mu* co-exists with the maize genome, and potential benefits/consequences of using *Mu* for insertional mutagenesis. We will focus particularly on how *Mu* elements may impact gene expression and how this may influence the use of these *Mu* stocks for generating loss of function alleles. Furthermore, we will introduce how we used *Mu*-insertion stocks in maize to perturb predicted gene-regulatory networks (GRNs) for functional validation *in vivo*.

Genetic properties of *Mu* in maize

Mu is a highly mutagenic plant transposon. The family of *Mu* TEs are only active in maize, but MULEs (*Mu*-like elements) are widespread in angiosperms [13]. *Mu*-induced mutations were first identified by Donald Robertson in 1978 when he described a maize line with a very high spontaneous mutation rate [19]. Upon selfing this line, many mutants were somatically unstable and had phenotypes with visible somatic sectors on the seed aleurone, which is characteristic of transposon excision or late somatic reversion events. Robertson referred to these lines with a high mutation rate as “mutator” active. When these “mutator” active lines were outcrossed to non-mutator active plants 90% of progeny maintained a high mutation rate/had “mutator” activity, which suggested this activity is inherited with non-Mendelian genetics. The “mutator” mutations in Robertson’s lines were caused by a novel class of transposons called *Mutator (Mu)* [14,18,26–29]. Robertson found that only when these *Mu* lines were selfed for many generations “mutator” activity could be lost [14]. It is now known that *Mu* transposons can be so

mutagenic because they prefer to transpose into low copy sequences or genic regions, have a high transposition frequency in the germline and germinal excisions or revertants are rare [17,21,22,30–33]. One of the reasons why *Mu* does not exhibit Mendelian genetics is because *Mu* is a multigenic trait or a two-component system with two types of elements, autonomous and non-autonomous, and many highly active stocks contain multiple autonomous elements [14,34].

Mu element family

DNA transposons can include both autonomous and non-autonomous elements. The *Mu* element family in maize consists of non-autonomous elements that have conserved ~220 bp terminal inverted repeats (TIRs), but each specific *Mu* element within the family has a unique internal sequence [13,14]. To date, there are 9 *Mu* elements that have been characterized with complete sequence (*Mu1*, *Mu2 /Mu1.7*, *Mu3*, *Mu4*, *Mu5*, *Mu6/7*, *Mu8*, *Mu9/MuDR*, *Mu13*) [28,29,35–37] and 3 with TIR sequence alone (*Mu10*, *Mu11* and *Mu12*) [38]. The ability for *Mu* elements to transpose is entirely dependent on the presence and activity of the autonomous *MuDR* element. *MuDR* elements encode two genes *mudrA* and *mudrB*: *mudrA* encodes the putative transposase-MURA and *mudrB* encodes a helper protein-MURB that may be required for integration of new insertions. The *Mu* element family is only active in maize because *MuDR*, with *mudrA* and *mudrB*, is only intact and active in the genus *Zea mays*. The putative mechanism for *Mu* transposition is described by *Mu* TIRs being bound and brought together by MURA to catalyze a DNA double-strand break and reintegrate *Mu* into the genome with the help of MURB [14].

Target site preferences

The genome-wide distribution of some DNA transposons provides evidence for preferential insertion into genic regions with specific chromatin, DNA methylation, and recombination activity [39–43]. *Mu* provides a clear example of preferential insertion near or into genes; however, the mechanism by which *Mu* targets genic regions remains unknown [14,22,41]. Unlike another DNA transposon family in maize, Ac/Ds (Activator/Dissociator), *Mu* inserts into regions that are genetically unlinked to the donor loci (relatively non-specific) and prefers 5' UTR or promoter regions of genes. Ac/Ds elements more frequently insert into exon or intron regions, but new insertions are usually linked to the donor loci [38,41]. *Mu/MuDR* and Ac/Ds are the most widely used

elements to develop transposon insertion stocks for insertional mutagenesis in maize. Although Ac/Ds might result in recovery of more loss-of-function mutations due to insertion site preferences into exon regions of genes, *Mu* stocks are more widely used because many more insertions in or near genes can be recovered [22,41].

Mu-suppressible alleles and an outward-reading promoter in *Mu*

The activity of *Mu* or ability to actively transpose can be described by two states: *Mu*-active and *Mu*-inactive. The *Mu*-active state is characterized by expression of *MuDR* and active transposition of *Mu* elements. Loss of *MuDR* activity and subsequent lack of transposition describes the *Mu*-inactive state [44–46]. The *Mu*-inactive state can occur either through genetic loss or segregation of *MuDR* or epigenetic silencing of *MuDR* [22,44,47]. It has been discovered that some *Mu*-induced mutations and phenotypes can be suppressed, but this suppression is dependent on the state of *Mu* activity. When a *Mu*-induced mutation is suppressed, this allele is referred to as a *Mu*-suppressible allele. *Mu* suppression occurs when in the absence of *Mu* activity (*MuDR* activity) the phenotype caused by the *Mu*-induced mutation reverts to that of wild-type. In contrast, when *Mu* activity is present the *Mu*-induced mutant phenotype is observed [48–50]. *Mu*-suppressible alleles have been described for several mutant alleles in maize, including *hcf106-mum1* [49], *Les28* [47], *a1-mum2* [51,52], *rs1* and *lg3* [53], *kn1* [54], and *rf2a* [55]. Suppression of these *Mu*-induced mutations resulted from both alterations in gene transcript structure and/or abundance and mRNA processing.

The molecular basis for *Mu*-suppressible alleles was first described for *Mu*-induced mutations in *High Chlorophyll Fluorescence 106* (*HCF106*), a gene in maize required for chloroplast membrane biogenesis [48–50]. The *hcf106-mum1* mutant allele is a recessive loss-of-function mutation caused by insertion of a *Mu1* element in the promoter of *HCF106* [48]. Homozygous *hcf106-mum1* maize seedlings are non-photosynthetic due to a defect in the chloroplast electron transport chain and exhibit a pale green phenotype which is lethal 3 weeks post-germination [48,50,56]. Barkan and Martienessen 1991 discovered that the *hcf106-mum1* mutant phenotype is expressed only in the *Mu*-active state. In contrast, when *Mu* is inactive, plants homozygous for *hcf106-mum1* exhibit a normal phenotype and the mutant phenotype is suppressed. The mutant phenotype is suppressed when *Mu* becomes inactive due to the activation of a promoter near the

termini of *Mu* that can direct transcription outward into the adjacent gene, an active outward-reading promoter [49]. In the *Mu*-inactive state, the *hcf106:Mu* mutant promoter could substitute for the *HCF106* wild-type promoter by generating transcripts encoding the entire wild-type ORF. To determine if the *Mu* promoter and the wild-type promoter are regulated similarly, the condition- and tissue-specific expression level of the *hcf106:Mu* promoter transcript in a *Mu*-inactive background was compared to that of the *HCF106*:promoter [49]. *HCF106* gene expression was known to be light-induced and greater in immature cobs compared to roots. Transcripts from both the *hcf106*:*Mu* promoter and *HCF106*:promoter were more abundant by a similar fold change after 24 hours of light exposure in leaf tissue and under normal conditions in immature cobs relative to roots. However, the *HCF106*:promoter can direct higher levels of transcript abundance than that of the *hcf106:Mu* promoter in leaf tissue under normal conditions [49]. In summary, these results demonstrate that regulation of the *Mu* promoter resembles that of the normal promoter, which suggests there is crosstalk between the *Mu* promoter and regulatory signals that control expression of the normal gene promoter .

Previously *Mu* suppression was thought to be the result of *Mu* inserting near the 5' of the gene, which would enable transcription initiation in *Mu* to read outwards and encode the entire ORF. However, *Mu* suppression has been described by other mechanisms for promoter, 5' UTR, intron and 3' UTR insertions [53,55,57]. Evidence of the activity of an outward-reading promoter that can mimic normal gene expression has only been reported for 5' UTR and promoter insertions. The mechanism of *Mu* suppression is still not fully understood, but it has been proposed to be a result of *Mu* promoter activity when there is not a transcriptional block from proteins (potentially transposases) bound to the ends of the *Mu* element in the *Mu*-active state [49].

Epigenetic regulation of *Mu* activity

There is consistent correlation of *Mu* elements becoming methylated when *Mu* activity is lost [47,58–60]. Methylation of active *MuDR* elements is associated with transcriptional silencing of *MuDR*, which leads to the absence of *Mu* transposition (i.e., no new *Mu* insertions). Due to the dependence of *Mu* element transposition on *MuDR* activity, in the absence of transposase (encoded by *MuDR*) there is often methylation of non-autonomous *Mu* elements. The presence of DNA methylation in non-autonomous

elements is reversible (lost) when *MuDR* elements are introduced through genetic crosses. In contrast, *MuDR* methylation appears to provide stable epigenetic silencing, even when non-silenced *MuDR* elements are introduced through crossing [13,15,25,61,62]. Like non-autonomous *Mu* elements that exhibit heritable, reversible inactivation, transposons may favor existing in highly methylated states as a mechanism for long-term survival [34,63]. This allows the transposon to limit the potential cost to the host via creation of deleterious alleles but still retain the potential to mobilize under certain conditions.

Characteristics of *Mu* stocks in maize

Creation and maintenance of the UniformMu population

Mu has been exploited for mutagenesis in maize to generate gene knockouts due to the high germinal mutation frequency. The *Mu*-indexed stock that is most widely used for insertional mutagenesis is the UniformMu population [22,64]. This population was created by introducing an active *MuDR* element into a color-converted W22 [65] line. An insertion of the non-autonomous *Mu1* element into the *Bz1* gene (*bz1-mum9* allele), a gene that is necessary to produce a purple anthocyanin pigment in the seed aleurone of maize, can be used as a genetic marker for the activity of *MuDR*. In the absence of *MuDR* activity, *Mu1* cannot excise from the *Bz1* gene and homozygous *bz1-mum9* alleles have uniform bronze colored seed aleurones. When *MuDR* is active, *Mu1* transposition is activated in somatic tissues and revertant sectors are purple, which results in a spotted aleurone phenotype. To ensure UniformMu stocks contain new insertions of *Mu* that are germinal and not somatic, loss of *Mu* activity is achieved via segregation of *MuDR* and is monitored by absence of the purple spotting phenotype [15,22,66,67].

To precisely map where transposition has occurred, *Mu* terminal inverted repeats (TIRs) can be used to tag or sequence-index new insertions relatively easily [66]. *Mu* TIRs are highly conserved ~220 bp repeats that are present at both ends of the element and read outwards into flanking genomic sequence. A TIR-specific primer can be used to amplify and sequence genomic DNA flanking new insertion sites in UniformMu mutant lines. These flanking sequence tags (FSTs) can then be mapped back to the genomic sequence of the W22 UniformMu inbred background to index each mutant to a specific locus [64]. High-throughput methods, such as Mu-seq, have been developed to enrich for

tagging of active and novel *Mu* insertions that are much less abundant than endogenous *Mu*-related sequences [68].

Other considerations for using UniformMu for insertional mutagenesis in maize

Properties of *Mu* that allow for the creation of *Mu*-indexed stocks with many new germinal insertions can also make it difficult to isolate an individual homozygous mutant allele. Each stock contains a variable number (usually 5-10, but sometimes more) of novel unlinked *Mu* insertions and is maintained by a sibling mating strategy to preserve deleterious mutations in a heterozygous state. This means researchers must confirm both the presence and zygosity of the allele and there is no guarantee the stock will contain the indexed *Mu* insertion. If the *Mu* insertion allele is present, it is ideal to backcross each line to standard wild-type W22 to reduce the likelihood that another unlinked *Mu* insertion is homozygous [66]. Another strategy to account for effects from genetic backgrounds with a high mutation load is to recover two independent mutant alleles in the gene of interest; however, this is not always possible.

Insertion alleles are isolated from the UniformMu population to disrupt gene function; however, the functional effects of the insertion may be suppressed in a *Mu* inactive background. One potential mechanism of suppression is transcriptional activity from an outward-reading promoter in *Mu*. Evidence suggests that over half of the *Mu*-induced mutations in the UniformMu population are suppressed in the absence of *Mu* activity [69]. This finding is further complicated by the target site preference of *Mu* into promoter or 5' UTR regions of genes [38,41], which increases the likelihood of an outward-reading promoter initiating transcription of the original ORF. Researchers should use caution when utilizing UniformMu insertion alleles and assuming knockouts without characterizing the transcripts for the mutant allele.

***Mu* mutagenesis in maize for perturbing inferred gene-regulatory networks**

In eukaryotes, gene regulatory networks (GRNs) describe the interactions between transcription factors (TFs) and the *cis*-regulatory elements (CREs) they recognize that regulate spatial and temporal expression of a portion of genes in the genome [70,71]. GRNs include a discrete number of TFs that transcriptionally regulate a much larger number of genes in the genome. Topological maps of GRNs are primarily generated to connect

trans-acting regulators (primarily TFs) to the target genes they regulate. These protein-DNA interactions can be inferred by various methods and many different types of data. One approach to generate GRNs is to use genome-wide expression datasets (RNA-seq) to cluster genes together based on their expression profiles. This method makes a guilt-by-association assumption that genes with shared expression patterns (co-expressed genes) are regulated by the same mechanism. To assign a direction of regulation where a TF is a regulator of a set of target genes, gene annotation information is required prior to clustering [70,72]. Another way to infer GRNs is to experimentally map TFs to target genes by methods that can be described as either TF-centered or gene-centered [70,73]. TF-centered approaches aim to identify a set of target genes for a TF of interest. Two TF-centered approaches are chromatin immunoprecipitation (ChIP-seq) and DNA-affinity purification (DAP-seq) [74]. ChIP-seq can identify indirect and direct target genes of a TF *in-vivo* but relies on the development of an antibody to the TF of interest. Alternatively, with DAP-seq, the TF of interest is transcribed and translated *in-vitro*, but only directly bound targets can be identified. Gene-centered approaches identify putative TF regulators for a set of target genes. One gene-centered approach to inform GRN generation is a Yeast-One-Hybrid (Y1H) screen. A Y1H identifies which TFs directly bind to CREs or promoters of a gene. These screens are limited in the number of interactions that can be tested and any resulting protein-DNA interactions identified may not be functionally relevant in maize [70,72,73]. Despite the limitations of each of these methods: co-expression, ChIP-seq, DAP-seq and Y1H, generating GRN predictions is advantageous to reduce the number of potential biologically relevant TF-target interactions from all possible combinations [70,72,75,76].

Due to the potential of rewiring GRNs to accelerate trait discovery and generate novel phenotypes, much effort has been devoted to identifying the regulatory components and topology of GRNs in crop species. While protein-DNA interactions predicted from these putative GRNs are useful for dimensionality reduction, these interactions must be validated *in vivo* to determine the prediction accuracy and functional relevance [75–78]. One method to evaluate GRN predictions is to generate knockouts for the predicted TF regulators and test expression of predicted targets [76,78,79]. In maize, *Mu*-indexed stocks can be used to isolate these TF loss-of-function alleles. While validation of inferred GRNs with single TF gene knockouts is a straightforward approach to connect TFs to

target genes, it is often necessary to study higher order mutants of multiple redundant TFs [75,80]. This is particularly important in maize due to the preferential retention of TFs following a recent whole genome duplication and the relatively large size of TF families [81–83]. Retention of TFs may be evolutionarily advantageous to coordinate gene expression; however, functional TF redundancy can be problematic if the effect of a TF knockout is masked by the compensation of another TF [82,84].

Concluding remarks

Our ability to understand how transposons are regulated and influence nearby gene expression will be important for future trait discovery in maize. The *Mu* family of transposons in maize has been utilized to study gene function and transposon biology. Continued research will be necessary to understand how *Mu* is able to provide a promoter that can interact with endogenous CREs to mimic expression patterns of different genes.

CHAPTER II. ABSTRACT

The highly active family of *Mutator* (*Mu*) DNA transposons has been widely used in maize for forward and reverse genetics. Prior studies have characterized examples of *Mu*-suppressible alleles which result in conditional phenotypic effects based on the activity of *Mu*. Phenotypes from these *Mu*-induced mutations are observed (i.e., not suppressed) in *Mu*-active genetic backgrounds, but absent when *Mu* activity is lost (*Mu*-inactive). For some *Mu*-suppressible alleles, phenotypic suppression likely results from an outward-reading promoter within *Mu* that is only active when the *Mu* element is silenced. We isolated 35 *Mu* insertion alleles from the UniformMu population that represent 24 different genes. Most of these alleles were isolated from insertions within gene coding sequences, but several 5' UTR and intron insertions were included. RNA-seq and *de novo* transcript assembly were utilized to document the transcripts produced from these *Mu* insertion alleles. For 20 of the 35 alleles there was evidence of transcripts initiating within the *Mu* sequence that read-through to the gene transcription termination site. This outward-reading promoter activity was detected in multiple types of *Mu* elements and does not depend on the orientation of the *Mu* insertion. Expression analyses of *Mu* initiated transcripts revealed the *Mu* promoter often provides gene expression levels and patterns that are quite similar to the wild-type gene. These results suggest the *Mu* promoter may represent a minimal promoter that can respond to gene *cis*-regulatory elements. Findings from this study have implications for maize researchers using the UniformMu population to isolate loss-of-function alleles, and more broadly highlights a strategy for transposons to co-exist with their host.

Chapter II entitled '*Mutator* transposon insertions often provide a novel promoter' is written and formatted as a manuscript. Some tables have been adapted and reformatted for thesis guidelines.

Erika L. Magnusson, Peng Zhou, Peter Hermanson, Yi-Hsuan Chu, Andrew C. Read, Candice N. Hirsch, Erich Grotewold, Nathan M. Springer (est. 2023).

Several authors have made contributions to Chapter II work. Erika L. Magnusson and Nathan M. Springer designed the experiments, interpreted the results, and completed the majority of the writing. Candice N. Hirsch, Peng Zhou and Andrew C. Read provided computational assistance for RNA-seq data processing. Peng Zhou helped isolate some of the mutants analyzed. Yi-Hsuan Chu and Erich Grotewold contributed RNA-seq data for a few mutant alleles. Candice N. Hirsch, Peng Zhou and Erich Grotewold interpreted the results of this manuscript and reviewed the writing. Peter Hermanson helped generate all the data for this manuscript. Chase Dickson contributed to supplemental data analysis.

CHAPTER II

Mutator transposon insertions often provide a novel promoter

Introduction

Transposon insertion stocks have been developed and successfully used to study gene function in several organisms, including plants [9–12], invertebrate animal models [85–88], bacteria [89], and a variety of single-cell eukaryotes [90,91]. Across species, the applicability of these transposon stocks vary due to properties of the transposon/transposase and the target genome, including sufficient transpositional activity, endogenous transposon copy number, transposon element type, family and size, integration site preference, and chromatin landscape of the genome [92]. In maize, there are two widely utilized DNA transposon families with sequence-indexed libraries: *Activator/Dissociation (Ac/Ds)* [41,93] and *Mutator (Mu)* [22]. Multiple *Mu* transposon populations have been generated in maize: UniformMu [22], BonnMu [23], Mu-Illumina [24], Pioneer Hi-Bred International's Trait Utility System for Corn (TUSC) [94,95], and Maize-targeted mutagenesis population (MTM) [69]. Both *Mu* and *Ds* elements preferentially transpose into low copy sequences or genic regions which is useful for mutagenesis [22,32,96–98]. However, the utility of *Ac/Ds* stocks for mutagenesis remains limited due to low copy number, low germinal insertion frequency, and the tendency for new copies to insert into sites genetically linked to the donor loci [41]. By contrast, *Mu* stocks are more widely used for reverse genetics because *Mu* has a high germinal insertion frequency, high forward mutation rate, and frequently inserts into genic regions that are unlinked to the donor loci [21,22,30,32,41,44].

Mu transposable elements are the most mutagenic plant transposons known, due to their high transposition frequency and tendency to insert into low copy sequences or genic regions [21,22,30–32]. The original *Mu* element was first described in maize in 1978 by Donald Robertson [19,99]. Robertson identified a maize line with what he called "mutator" activity that had a very high forward mutation rate, 50- to 100-fold greater than that of the background. Upon outcrossing mutator plants to non-mutator plants, 90% of the progeny retained the high mutation frequency which suggested there is non-Mendelian inheritance of "mutator" activity. It is now known that *Mutator* (*Mu*) transposons were responsible for the mutations in Robertson's lines [14,26–29]. There is evidence that *Mu* elements in maize can reach high copy numbers because they can transpose at

frequencies of 100% [33,100], have a high germline mutation rate, and are rarely excised from the germline (excision rate < 10^{-4}) [16,17,20,101–103].

The *Mutator* system is a two-component system with one autonomous element, *MuDR*, and several non-autonomous elements (*Mu1* to *Mu13*) [31,37]. Non-autonomous *Mu* elements all contain similar ~220 bp terminal inverted repeats (TIRs), but each class of element has unique internal sequences [14,18]. *Mutator* activity is dependent on the presence of an active autonomous *MuDR* element to encode the proteins necessary for transposition of itself and non-autonomous elements [51,104]. *Mu*, like *Ds*, preferentially inserts into gene rich regions; however, *Mu* exhibits a stronger preference than *Ds* for 5' UTR or promoter regions [38,41]. These genic regions where *Mu* lands have distinct chromatin, DNA methylation, and recombination activity that could influence *Mu* element targeting [14,105]. *Mutator* activity can be epigenetically regulated, such that some plants with *MuDR* are in fact *Mu*-inactive due to heterochromatin-mediated silencing of the *MuDR* coding sequences, which is accompanied by high levels of DNA methylation [22,51,58]. Along with epigenetically silenced *MuDR*, the non-autonomous *Mu* elements do not exhibit evidence for transposition and are methylated in the *Mu*-inactive state. In the *Mu*-active state the autonomous and non-autonomous *Mu* elements are hypomethylated and products from *MuDR* can mobilize *Mu* elements [44,45]. To develop UniformMu populations in maize that are genetically stable (in *Mu*-inactive genetic backgrounds) new germinal *Mu* insertions from lines with *MuDR* activity (*Mu*-active) are stabilized by selecting against somatic transposition of *Mu* using the *bronze1-mum9* (*bz1-mum9*) mutation as a genetic marker for *MuDR* activity [22,64,66,101,106].

Mutator has been used to mutagenize many genes and isolate loss-of-function alleles [38,64,107–110]. In many cases these *Mu* insertion alleles produce a stable phenotype that does not change depending upon the epigenetic state of *Mu*. However, there is evidence that the phenotypic consequences for some of these *Mu*-induced alleles can be suppressed depending on the state of *Mu*. *Mu*-induced mutations that are suppressible, *Mu*-suppressible alleles, exhibit a mutant phenotype in *Mu*-active genetic backgrounds that can be suppressed, returning to a wild-type phenotype, when *Mu* activity is lost (*Mu*-inactive). *Mu*-suppressible alleles were well characterized in a recessive loss-of-function mutation, *hcf106-mum1*, caused by insertion of *Mu1* in the promoter of *HCF106*, a gene

in maize required for chloroplast membrane biogenesis [48–50]. Homozygous *hcf106-mum1* maize seedlings expressed a non-photosynthetic, pale green mutant phenotype only in the absence of *Mu* activity (*Mu*-inactive) [48–50,56]. It was found that in *Mu*-inactive stocks an outward-reading promoter near the termini of *Mu* can direct transcription outward into the adjacent gene and substitute for the *HCF106* promoter [49]. Since this discovery several other Mu-suppressible alleles have been described, including *Les28* [47], *a1-mum2* [51,52], *rs1* and *lg3* [53], *kn1* [54], and *rf2a* [55]. The full extent and molecular mechanisms of suppressibility of *Mu*-induced mutations have not been characterized widely but may be a property of the *Mutator* system [14].

We sought to further characterize the frequency of promoter activity in *Mu*-induced mutations and properties influencing the promoter's ability to direct transcription outward has not been previously reported. Here we characterized the transcripts of 35 *Mu* insertion alleles. We find evidence that many (n = 20) of these alleles result in the production of two transcripts: one initiating at the normal gene promoter and another initiating from a *Mu* outward-reading promoter. This *Mu* outward-reading promoter appears to be functional in several of the non-autonomous *Mu* elements and is not dependent upon *Mu* orientation. Interestingly, our findings suggest that the *Mu* promoter is a minimal promoter that often shows expression levels and patterns quite similar to the gene it is inserted within. These findings highlight a potential strategy for co-evolutionary interactions between transposons and their host genomes.

Results

Characterization of transcripts arising from genes with *Mutator* insertions

To investigate the effect of *Mutator (Mu)* insertions on transcript structure we isolated homozygous mutants for 35 insertions in 24 genes [22,66]. These included 9 insertions in 5' UTR sequence, 22 insertions in coding regions, and 4 insertions in introns (Figure 1, Table S1). These frequencies do not necessarily reflect the spectrum of insertion sites for all *Mu* elements. We focused on selection of insertions within coding sequences as the mutants were originally selected as putative loss-of-function alleles for maize transcription factors. We generated RNA-seq data for three biological replicates of each homozygous mutant and wild-type allele. A single tissue for each mutant allele was selected to generate RNA-seq data based on evidence of wild-type allele expression

(Table S2). The expression level of the mutant allele was documented by aligning RNA-seq data to the W22 reference genome [65] and only 8/35 alleles exhibited significantly lower transcript abundance in the mutant relative to the wild-type (Table S3). The majority, 25/35, of mutant alleles do not have any significant change in transcript abundance relative to wild-type, and 2 alleles have significantly higher transcript abundance (Table S3). Assessment of the mapped transcript reads derived from homozygous mutant plants revealed reduced coverage at the site of the *Mu* insertion (Figure S1). The drop in coverage flanking the *Mu* insertion site is expected if there is a novel junction and/or sequence present at this region in the mutant allele transcript relative to the W22 reference genome.

There are several potential transcript structures that might be expected to be produced from *Mu* insertion alleles in comparison to the full-length wild-type W22 gene transcripts (Figure 2A). Mutant allele transcripts could include: read-through transcription resulting in retention of the full *Mu* sequence (*Mu* read-through transcript), novel splicing events that include retention of a portion of the *Mu* sequence (*Mu* spliced transcript), transcript initiation at the normal gene transcription start site–TSS with premature termination in the *Mu* sequence (gene promoter-*Mu*), or transcript initiation from an *Mu* outward-reading promoter reading through to the normal termination site (*Mu* promoter transcript) (Figure 2A). These potential *Mu* insertion allele transcript structures are not necessarily mutually exclusive. To determine if *Mu* insertion alleles produced read-through transcripts with all or a portion of *Mu* sequence retained (*Mu* read-through or *Mu* spliced transcripts) we performed RT-PCR with gene-specific primers that flank the *Mu* insertion site (Figure 2B). Although we expected to amplify a variable sized product in the mutant alleles relative to wild-type, we were not able to amplify products with primers flanking the *Mu* insertion site for 7 mutant alleles tested (see examples in Figure 2C). This suggests that read-through transcription of the *Mu* insertion with retention of partial or complete *Mu* sequences in the mRNA is unlikely or rarer.

To investigate transcript structure of the *Mu* insertion alleles, we generated *de novo* transcript assemblies for each mutant (35 alleles for 24 genes) and wild-type W22 RNA-seq (5 tissues) dataset for a total of 40 transcriptome assemblies. The *de novo* transcriptome assemblies for the W22 control samples generated full-length assemblies

for 21 of the 24 genes in the respective tissue selected to sample for RNA-seq (Table S3). Two of the remaining three genes, *WRKY8* and *WRKY2*, could be assembled as two overlapping W22 transcripts due to a gap in coverage in one region. The other gene, *HSF6*, lacked adequate coverage to assemble any mutant, *hsf6-m1* and *hsf6-m2*, or control transcripts and was removed from subsequent analyses. In total, transcript assemblies for 33 mutant alleles that aligned to the respective wild-type gene with a *Mu* insertion allele (23 genes) were identified and further characterized (Table S3). The transcripts from the mutant alleles could include transcripts arising from the gene that are 5' and 3' of the *Mu* insertion or exclusively 5' or 3' of *Mu* (Figure 3A).

For 20 alleles we identified an assembled transcript that matched gene sequences 5' of the *Mu* insertion site (Figure 3A, Table S1, Data File S1). Most (17/20) mutant alleles with assembled transcripts that contain gene sequence 5' of *Mu* include at least a portion of *Mu* sequence at the 3' end which indicates the transcript reads into the *Mu* element and terminates. The 13 alleles that did not have transcripts assembled 5' of *Mu* reflect insertions very near the 5' end of the gene (*Mu* insertion sites within the first 33% of gene cDNA). In these cases, it is likely that if a short transcript (< 200 bp) was produced by the normal gene promoter it would be underrepresented in the RNA-seq data due to the size selection step during library preparation [111] and the lack of an assembly could reflect technical bias against short transcripts rather than absence of this transcript. The majority (31/33) of mutant alleles produce transcripts that include sequences 3' of the *Mu* insertion site. Many (21/31) of these transcripts that contain sequences 3' of the *Mu* insertion site contain a portion of recognizable *Mu* sequence at the 5' end (Figure 3A, Table S3). The remaining 10 alleles have partial transcript coverage 3' of the *Mu* insertion and often (7/10 alleles) have relatively low expression levels (< 5.3 FPKM). Examples of assembled transcript structures observed for two *Mu* alleles, *jmj13-m4* and *sbp20-m2*, compared to the W22 allele are shown in Figure 3B. These mutant transcript assemblies suggest that many of the *Mu* insertions within the coding sequence result in the presence of two partial transcripts: a transcript initiating within the gene promoter and terminating within *Mu* or prematurely terminating (gene promoter-*Mu* or gene-promoter partial transcripts) and a transcript initiating at a *Mu* outward-reading promoter reading through the end of the gene (*Mu* promoter transcript).

Mu promoter initiation does not strictly depend on a specific *Mu* element or orientation

There are multiple distinct members of the *Mutator* transposon family that could be mobilized, and specific insertions could occur in forward or reverse orientations relative to the gene sequence. To further understand the impact of specific *Mu* transposons and their orientation upon the potential of the Mu *promoter* to initiate outward-reading transcripts, we sought to characterize the identity and orientation of each mutant allele *Mu* insertion. The *Mu* sequence from each *de novo* assembled transcript, either found at the 3' end of transcripts with gene sequence 5' of Mu (gene promoter partial transcripts) or 5' end of transcripts with gene sequence 3' of Mu (Mu promoter transcripts), was used to perform a BLAST search against representative examples of *Zea mays Mu* elements. *Mu* element identity was predicted based on the *Mu* element with the greatest similarity to *Mu* sequence from each transcript (Figure 4). Most of the assembled transcripts with *Mu* sequence (24/32 tested for 21 alleles) only contain *Mu* sequences that align to the *Mu* terminal inverted repeats (TIRs) with only a subset that include internal *Mu* sequences (Figure 4). The transcript assembly *Mu* sequence alignments suggest that there are 8 *Mu1* or *Mu1.7* elements (these cannot be separated based on the TIR regions alone), 2 *Mu3*, 5 *rcy:Mu7*, 5 *Mu8* and 1 *Mu13* element in our set of alleles. The 5' and 3' TIRs of *Mu* elements often have some polymorphisms between the two TIRs and these could be used to predict the orientation of the *Mu* insertion for 15 of the 21 alleles with the presence of *Mu* sequence in their transcript assembly. One mutant allele with *Mu* inserted into an intron, *bzip76-m2*, had evidence for only internal *Mu* sequences, potentially suggesting that this allele may produce a *Mu* spliced transcript instead of a *Mu* promoter transcript. The predictions for *Mu* element identity and insertion orientation based on sequence alignments were tested using outward-reading PCR primers with specificity to either the 5' or 3' internal *Mu* sequence (Figure S2). We were able to confirm the identity for 14 of the 15 *Mu* insertions that had alignment-based predictions and for 12/14 of these the predicted orientation was supported by PCR-based testing (Figure S2, Table S4). The *Mu* element identity and orientation for the remaining alleles without discernible predictions for *Mu* orientation based on transcript alignments was determined using *Mu* element specific primers (Figure S2).

Evidence for transcript termination and initiation within *Mu*

The observation that the majority of *Mu* insertion alleles within genes generate two separate transcripts could reflect the failure to assemble *de novo* transcripts through the *Mu* element or the presence of two independent transcripts, one initiating at the gene promoter and the other initiating at an outward-reading promoter within *Mu*. Our previous RT-PCR results (Figure 2C) suggest that at least a portion of *Mu* is not retained in a full-length transcript; however, these results could also reflect difficulty in successfully amplifying through the full *Mu* element. To rule out the potential of *Mu* insertion alleles producing transcripts that read through the *Mu* element, several approaches were utilized. One approach we used was to perform RT-PCR using gene-specific primers and primers within the predicted *Mu* element that flank the *Mu* sequence observed in the *de novo* assembled transcripts (Figure 5A, Table S5). To ensure our RT-PCR *Mu* sequence amplification approach would be comparable to the transcriptome assembly results, mutant and wild-type allele cDNA was generated from the same tissue type sampled for RNA-seq. This approach would allow us to determine how much of the *Mu* element was retained in each mutant allele transcript tested.

Mutant transcript boundaries were defined by the presence of RT-PCR amplification from a gene specific primer and a primer with specificity to *Mu* sequence. Resulting amplicons are categorized based on the *Mu* sequence amplified relative to that observed in the *de novo* assembled transcript: within; *Mu* sequence observed in the assembly, slightly outside; *Mu* predicted element sequence ≤ 200 bp more internal than that included in the assembly, or further into; *Mu* predicted element sequence > 200 bp more internal than that included in the assembly. In all cases we were able to successfully amplify regions that were within the assembled transcripts (Figure 5B). For several of the alleles we were also able to successfully amplify RT-PCR products using *Mu* primers that were slightly outside the assembled *Mu* sequence. However, we failed to get amplification using primers further into the *Mu* sequence (more internal) for all transcripts tested which suggests that this *Mu* sequence is not present in the mature transcript (Figure 5B, Table S5).

Transcripts from the mutant allele often have similar abundance to the wild-type allele

We were interested in comparing the expression level of each mutant transcript with the wild-type allele to document potential variability in transcript abundance. To compare expression levels between the mutant and wild-type alleles, we focused on RNA-seq reads that mapped to exon regions with shared sequence between the wild-type transcript and either the mutant gene promoter partial (terminated within *Mu* or prematurely terminated) or *Mu* promoter transcript (Figure 6A). The expression level of shared exon regions (referred to as CPM per fragment) between the mutant gene promoter partial and wild-type transcripts reveal highly similar transcript abundances, R^2 = 0.967 for 18 transcripts (Figure 6B). The *Mu* promoter transcript abundance was generally quite similar to the levels of the wild-type transcript, but *Mu* promoter transcripts exhibit more variation with an R^2 = 0.426 for 19 transcripts (Figure 6C). There were several examples of lower expression levels for the *Mu* promoter transcript relative to the wild-type transcript (Figure 6C). The observation that the *Mu* promoter transcript abundance is often similar to the wild-type in a single tissue suggests that this *Mu* outward-reading promoter may provide expression levels comparable to the normal gene promoter.

***Mu*-derived transcripts can maintain similar tissue-specific patterns**

We were also interested in assessing whether mutant transcripts derived from the gene or *Mu* promoter would exhibit similar patterns of expression across multiple tissues. To evaluate tissue-specific expression, we performed RT-qPCR by amplifying transcript regions that are shared between the wild-type and mutant gene promoter-*Mu* or *Mu* promoter transcripts (Figure 7A). To ensure we could test tissue-specific expression patterns, we selected genes with variable levels of wild-type gene expression across multiple tissues. Mutant gene promoter-*Mu* transcripts maintain relative expression levels that are very similar to wild-type transcripts across all tissues tested (Figure 7B) suggesting similar expression patterns for this mutant transcript and the wild-type transcript. The mutant *Mu* promoter initiated transcripts often maintain wild-type tissue-specificity but frequently have lower relative transcript abundance–higher Delta Ct (Figure 7B). The *Mu* element identity (*Mu1.7, Mu3, rcy:Mu7* or *Mu8*) and *Mu* insertion position (5’ UTR or CDS) vary among the ten mutant alleles tested by RT-qPCR. The

finding that all 10 mutant allele *Mu* promoter transcripts have patterns of expression similar to wild-type tissue-specificity suggests the ability of the *Mu* outward-reading promoter to mimic wild-type gene expression patterns is not entirely dependent on the specific *Mu* element or where *Mu* inserts within a gene.

Discussion

Previous work characterizing mutant alleles from insertions of *Mutator* transposons in maize genes identified the presence of an outward-reading promoter in *Mu* [49,112]. The ability of this *Mu* outward-reading promoter to initiate transcription is conditional upon the epigenetic state of *Mu* [49]. *Mutator* elements can be in either one of two states: an active state (*Mu*-active) where there is a high forward mutation rate from the presence of active *MuDR* transposons or an inactive state (*Mu*-inactive) without *MuDR* activity [51]. The state of *Mu* activity can be monitored by the extent of DNA methylation in sequences of *Mu* terminal inverted repeats (TIRs); plants with *Mu*-active state elements are marked by hypomethylation and plants with *Mu*-inactive exhibit hypermethylation [18,28,49,58,113]. There are several reports demonstrating that insertions of *Mu* transposons into maize genes can lead to mutations whose phenotypes are suppressed only in the absence of *Mu* activity (*Mu*-inactive) [50,51,53–57,108,114]. In the *Mu*-inactive state, the *Mu* promoter becomes active and initiates transcription directed outward into the adjacent gene to restore the phenotype of the *Mu*-induced allele to that of its progenitor [49]. Mutant alleles with phenotypes that depend on the activity of *Mu* for expression are known as *Mu*-suppressible alleles [48].

In our study, we provide evidence that transcript initiation in *Mu* and potential activity of a *Mu* outward-reading promoter is a common phenomenon for mutant alleles isolated from the UniformMu mutant population in maize. Most mutant alleles (20/33) isolated in our study have transcript assembly evidence of transcript initiation in *Mu* sequence. Although we did not directly examine *Mu* activity in these stocks, we can infer that these mutant alleles are likely in a *Mu*-inactive state as one step in the creation of the UniformMu population includes selection for kernels that lack evidence of *Mu* activity prior to identification of insertion alleles [22]. All previously reported *Mu*-suppressible alleles in maize with evidence of an active promoter in *Mu* were generated from a *Mu* insertion in the promoter or 5'UTR of a gene [49,53,55,57,112]. However, the previously

characterized *Mu*-suppressible alleles were all detected based on phenotypic effects and likely required production of transcripts from the *Mu* promoter that could produce a functional protein. *Mu* elements inserted near the normal transcription start site provide the potential for production of a transcript that can encode the full ORF of the gene. However, these insertions near the 5' end of the gene may also allow for activity of the outward-reading promoter within *Mu* based upon position of this *Mu* promoter very near the site of the gene promoter which may allow for the *cis*-regulatory elements of the gene to influence this *Mu* promoter.

Our results show that the activity of the outward-reading promoter in *Mu* is not dependent on the *Mu* insertion site being in the gene promoter or 5' UTR. For the 20 alleles with evidence of a *Mu* promoter initiated transcript, 13 were isolated in coding sequences from various positions spanning the gene length, 5 were within the 5' UTR, and 2 were within introns. These results suggest that the distance of the *Mu* promoter from the normal gene promoter does not necessarily determine activity of the *Mu* promoter. Previous reports indicate the *Mu* outward-reading promoter is located near the edge of the *Mu* element–potentially initiating transcripts from within the *Mu* TIR sequence [49,112]. When we aligned sequences from *Mu* promoter transcripts of 20 alleles to predicted *Mu* element sequences, the *Mu* transcribed sequence from 15/20 transcripts mapped entirely to the TIR sequence. Although the exact location of the promoter within *Mu* was not precisely defined, the *Mu* outward-reading promoter is likely located near the termini of *Mu*. We find examples of multiple Mu elements (*Mu1.7, Mu3, rcy:Mu7* and *Mu8*) and *Mu* insertions in either the forward or reverse orientation relative to the gene TSS can provide an outward-reading promoter. Several prior studies also suggested that *Mu*-suppressible alleles could include different types of *Mu* elements and orientations [53,108]

The UniformMu induced mutations are a widely used tool for functional genomics in maize [22,64,66,67,115]. While the silencing of *Mu* transposition is quite useful for ensuring that the detected *Mu* insertions represent germinal rather than somatic insertion events, it also has the potential to create *Mu*-suppressible alleles through potential activation of *Mu* outward-reading promoters. Our study provides evidence that for many UniformMu mutant alleles the *Mu* element provides an outward-reading promoter that can direct

transcription into adjacent gene sequences. *Mu* promoter activity seems to be similar to *Mu* suppression in terms of frequency as it has been reported that in the absence of *Mu* activity over half of the *Mu*-induced mutations are suppressed [69]. Although we found that *Mu* promoter activity does not depend on insertion site, it is known that *Mu* preferentially inserts into promoter and 5' UTR regions [38,41,65]. This means that many *Mu* insertion alleles within the UniformMu population represent 5' UTR insertion events. Researchers should use caution when interpreting the phenotypic results of these insertion events as it is possible that *Mu* promoter transcripts could complement the insertion. Failure to obtain RT-PCR products when using primers that flank the *Mu* insertion site does not necessarily indicate a true loss-of-function allele as a *Mu* initiated transcript could still be produced. *Mu* insertions into regions upstream of the gene coding sequence are not only the most abundant (42%) in the UniformMu population, but also allow the potential for the promoter in *Mu* to drive a transcript that includes the entire original ORF and complement the mutant phenotype.

The tissue-specificity of the *Mu* outward-reading promoter has not been well characterized. The ability of the *Mu* outward-reading promoter to suppress phenotypes of *Mu* insertion alleles suggests the ability to drive expression in tissues in which the gene product is needed to normally function. Prior work on *hcf106-mum1* suppressible alleles revealed that both the wild-type *Hcf106* and *hcf106:Mu* exhibit similar patterns of expression response to light [49], albeit with higher levels of expression for the wild-type gene compared to *hcf106:Mu*. Our collection of transcripts initiated within *Mu* provided an opportunity for a broader characterization of the activity of the *Mu* outward-reading promoter. Our findings suggest that the outward-reading promoter seems to be able to mimic the gene promoter in terms of expression level and tissue-specificity and has limited inherent patterns of expression. The RNA-seq data was generated from multiple tissues depending on which tissues exhibit high levels of expression for the wild-type gene. We found that the transcript abundance (CPM per fragment) of the mutant *Mu* promoter transcripts tend to be similar to wild-type for most alleles and that there were not consistent differences in the expression of the *Mu* promoter transcripts among the tissues. Several genes that were selected for characterization of expression patterns by RT-qPCR normally exhibit variation in expression among the profiled tissues. We found that the *Mu* transcripts tended to mimic these tissue-specific patterns. Although *Mu*

promoter transcripts follow wild-type tissue-specific patterns, they often exhibit expression levels that are slightly lower than that of gene promoter partial transcripts relative to wild-type. Our results, along with previous reports [46,49,54,55,57], imply that *Mu* might provide a minimal outward-reading promoter that can interact with genic *cis*-regulatory elements to condition expression patterns and levels that are similar to the wild-type gene.

There are several mechanisms by which maize transposons can minimize the functional impact of insertions into genic regions. The *Mu* family of transposons seems to have adopted a mechanism of providing an outward-reading promoter that is active when the *Mu* is silenced. The coupling of this mechanism with the preferential insertion within promoters and 5' UTRs provides the opportunity for *Mu* elements to insert within open chromatin regions while limiting potential deleterious consequences. Our findings and previous studies [49] suggest that the *Mu* promoter relies on interactions with genic *cis*-regulatory elements to mimic normal gene expression patterns. By providing a minimal promoter that can mimic the expression pattern of the gene *Mu* elements can potentially insert and increase in copy number with limited effects on the long-term survival in the host. This provides an elegant solution for a transposon to limit the consequences of its proliferation.

This study provides evidence that *Mu* transposon insertions often result in complex transcripts for the gene rather than clear loss-of-function alleles. Our mutant allele transcript assemblies frequently include examples of termination and initiation in *Mu* sequence. Transcripts initiating from *Mu* are likely derived from a *Mu* outward-reading promoter that may produce functional transcripts if these include the full ORF of the gene. These results have implications for the many researchers that utilize UniformMu for reverse genetics. Further studies are necessary to document whether the *Mu* outward-reading promoter requires a *Mu*-inactive state and to uncover the mechanisms that allow the *Mu* promoter to interact with genic *cis*-regulatory elements. The ability of *Mu* to provide an outward-reading promoter also has implications for future transposon biology. The system by which conditional activity of a *Mu* promoter determines whether *Mu* can suppress a mutant allele should be utilized to understand the relationship between transposons and host genomes.

Methods

Isolation of homozygous mutant alleles from the UniformMu population in maize

Transposon-indexed seed stocks were ordered from MaizeGDB Stock Center [22,116]. Most alleles were selected based on insertions into the coding sequence or 5' UTR. Seeds were planted in the field to maintain seed stocks. At the Vegetative 3 (V3) developmental stage leaf tissue was collected for DNA isolation. Mutant alleles were genotyped to identify the presence and zygosity of *Mu* with gene-specific primers flanking the *Mu* insertion and a primer with specificity to the *Mu* TIR regions: 9242, as described in [66]. Homozygous transmissible alleles were then isolated after backcrossing twice to the W22 r-g inbred, if possible, to reduce the original mutation load from the transposon-indexed stock (Table S1).

Plant material for RNA-seq samples

Wild-type tissue-specific expression data from B73v4 [117] and W22 [118] were used to identify tissues where each of the 24 maize genes with a *Mu*-insertion mutant allele had moderate to high expression. To capture expression of each of the 24 genes in both mutant and control conditions, 5 different tissues were selected to sample: coleoptile tip, seedling leaf, imbibed embryo, tassel, and tassel stem (Table S2). Three biological replicates of the mutant allele and at least three biological replicates of control W22 r-g were sampled for RNA-seq from the respective tissue selected for each gene (Table S1). Samples for each tissue were collected on the same day at the same time with the 24 h time sampled listed in parentheses. At anthesis, tassel stems (~3 cm) and whole tassels (anthers unextruded) were sampled (9:00) from three plants in the field each and pooled for one biological replicate. Embryos were dissected (from 8:00-11:00) after imbibing seeds in distilled water for 48 h at 31°C and 5 embryos were pooled for one biological replicate. For seedling leaf tissue, the V3 collared leaf was sampled (9:00) from each seedling 10 days after sowing (DAS) in 16 h light 28°C, 8 h dark 24°C growth chamber conditions and 3 leaves were pooled for each biological replicate. Coleoptile tips were sampled (9:00) from seeds 6 DAS in 30°C dark conditions using a paper towel cigar roll method for germination [119] and 3 tips (~2.5 cm) were pooled for each biological replicate.

RNA-seq data processing

Total RNA was extracted using the RNeasy Plant Mini Kit (QIAGEN, Cat # 74904), quantified internally and externally by University of Minnesota Genomics Center (UMGC) with the Quant-iT RiboGreen RNA Assay Kit (Thermo Fisher, Cat # R11490), and quality checked with the Agilent 2100 Bioanalyzer. One biological replicate of the mutant allele, *gras52-m1*, had low RNA quality and was discarded prior to sequencing. Sequence libraries were prepared from a minimum of 500 ng of total RNA using the standard TruSeq Stranded mRNA library protocol (Illumina, Cat # 20020595) and sequenced on the NovaSeq 6000 S4 flow cell to produce at least 20 million 150 bp paired-end reads for each sample. For all samples with paired-end sequencing, both library construction and sequencing were done at UMGC. Library construction and sequencing for two mutant alleles, *hsf24-m3* and *hsf24-m4*, and W22 control sampled from tassel tissue was done externally at the Genomic Core at Michigan State University. For these samples, libraries were prepared from 2 µg of total RNA using the TruSeq RNA Sample Prep Kit (Illumina, Cat # FC-122-1001) and sequenced on the HiSeq 4000 to produce at least 18 million 50 bp single-end reads.

For all samples, sequencing reads were then processed through the nf-core RNA-Seq pipeline [120,121] built with Nextflow v20.10.0 [121] for initial QC and raw read counting. Reads were trimmed using Trim Galore! v0.6.5 [122] and aligned to the W22 reference genome [65] using Hisat2 v2.1.0 [123] with default parameters ("hisat2 -x $db $input -p 12 --met-stderr --new-summary"). Uniquely aligned reads were counted per feature by featureCounts v2.0.1 [124]. Raw read counts were normalized by library size and corrected for library composition bias using the TMM normalization approach in edgeR v3.28.0 [125], to give CPMs (Counts Per Million reads) for each gene in each sample allowing direct comparison between mutant and control samples (Table S1). CPM values were normalized by gene CDS lengths to give FPKM (Fragments Per Kilobase of exon per Million reads) values (Table S1). Genes were considered expressed if their CPM was ≥ 1 in at least one sample per tissue.

Identification of differentially expressed genes

Raw read counts of expressed genes (CPM ≥ 1 in at least 1 sample per tissue) from all replicates of each mutant allele and W22 control from the same tissue were used to call

differentially expressed (DE) genes, false discovery rate [FDR] adjusted p-value < 0.05 and a minimum fold change of 2 (DESeq2 v1.30.1 [126]) (Table S3).

Transcriptome profiling

Reads from RNA-seq data of combined biological replicates for each allele, mutant or control, were *de novo* assembled into transcripts with TRINITY v2.5.1 [127]; Reads were quality trimmed with Trimmomatic v0.33 [128] and strand-specificity was not defined. A local blast database (SequenceServer [129]) was created for each *de novo* transcriptome assembly to identify transcripts aligning to the W22 gene cDNA in both the mutant and control. W22 control transcript assemblies for each gene were analyzed first by both BLASTn [130] and the ExPASy translate tool [131] to confirm TRINITY could assemble the full-length gene cDNA from the RNA-seq short-read data. The canonical ORF of each gene was identified by comparing the annotated W22 gene cDNA sequence to sequences of orthologous genes in other grass species (i.e., *Sorghum Bicolor, Setaria Italica, Oryza sativa*) via BLASTx. Mutant allele transcript assemblies were further analyzed by BLASTn and Expasy to determine the effect of the transposon insertion and identify if *Mu* sequence was transcribed (Table S3).

Predictions of *Mu* element identity and orientation

Sequence from each mutant assembled transcript was used as a query against all public sequencing databases–NCBI to identify if there were any hits to *Mu* elements. Complete sequences of representative *Zea mays Mu* elements with transcript assembly hits: *Mu1* (X00913.1), *Mu1.7* (Y00603.1), *Mu3* (JX843286.1:132-1963), *Mu4* (X14224.1), *Mu5* (X14225.1), *rcy:Mu7* (X15872.1), *Mu8* (X53604.1), Mu13 (HQ698272.1), *Mu17* (HQ698276.1), and *MuDR-MudrA* and *MudrB* (M76978.1), were used to create a local *Mu* element BLAST database (SequenceServer [129]) (GenBank Nucleotide Accessions from NCBI). *Mu* sequence from each assembled transcript was then BLAST against only *Mu* element sequence and top hits were used to predict the *Mu* element for each allele. The *Mu* sequence from assembled transcripts of each mutant allele was then aligned to the complete sequence of the predicted *Mu* element. *Mu* element insertion orientation could be predicted when transcribed *Mu* sequence either only aligned to or aligned with greater similarity to 5’ or 3’ regions of *Mu*.

PCR confirmation of *Mu* element identity and orientation

For each predicted *Mu* element, outward-reading primers with specificity to either the 5' or 3' sequence of *Mu* were designed (Table S4). We refer to a forward orientation of *Mu* when *Mu* 5' TIR sequence relative to 3' TIR sequence is closest to the gene TSS and reverse orientation when Mu 3' TIR sequence is closest to the gene TSS. PCR was performed on mutant allele gDNA with specific combinations of gene-specific (referred to as F and R) and *Mu*-specific primers (referred to as 5 and 3) to confirm the identity and orientation of *Mu* (Table S4). The presence of amplicons from F:5 and/or R:3 indicates a *Mu* element with forward orientation while amplification from F:3 and/or R:5 indicates reverse orientation. The *Mu* primers designed had specificity to 5' or 3' sequences of a specific *Mu* element. Amplification of gDNA using these *Mu* element-specific primers was considered *Mu* element identity confirmation.

Mutant assembled transcript structure assessed by RT-PCR

Mu sequences from transcript assemblies of all mutant alleles with a shared *Mu* element identity were aligned to the complete sequence of that *Mu* element. Outward-reading PCR primers were designed with specificity to regions of the *Mu* sequence included in the transcript assembly and regions outside of the assembly for each mutant allele (Table S5). Gene-specific primers flanking the *Mu* insertion were designed with specificity to both mutant allele gDNA and cDNA sequence (Table S5). For each mutant allele, PCR was performed on both gDNA and cDNA to determine if mutant transcripts terminated and initiated in *Mu* at the predicted transcript assembly sites. As a control, PCR was first performed on mutant allele gDNA to detect presence of amplification from each primer-set designed with specificity to gene and *Mu* sequence included in and outside of the transcript assembly. Combinations of gene-specific and *Mu*-specific primers that amplified mutant gDNA were then used to test for presence of amplification from mutant cDNA (Table S5). To test the predicted transcript structures for six mutant alleles by RT-PCR, total RNA was extracted from the same tissue type sampled for RNA-seq. Extracting RNA from the same tissue type RNA-seq was performed on allows for direct comparison between our RT-PCR results and the assembly results without bias of amplification from tissue-specific isoforms. Tissue was sampled for at least three biological replicates of each mutant allele and W22 control using tissue sampling methods listed above in the Plant Material section. Total RNA was extracted from ~100

mg of tissue/sample using TRIzol™ Reagent (Thermo Fisher, Cat # 15596026), DNase treated (TURBO DNA-free™ Kit, Thermo Fisher, Cat # AM1907), and quantified with Quant-iT RiboGreen RNA Assay Kit (Thermo Fisher, Cat # R11490). RT-PCR reactions for each primer set were performed on RNA (50-75 ng/ul) from at least two biological replicates of the mutant allele and control (QIAGEN OneStep RT-PCR Kit, QIAGEN Inc., Cat # 210212). The bounds of where transcriptional termination and initiation occurs within *Mu* for each mutant allele was determined by presence of amplification from gDNA and absence of amplification from cDNA. If there was amplification of mutant cDNA from primers designed to amplify regions outside of the transcript assembly, then more *Mu* sequence was transcribed than the assembly predicted.

Transcript abundance (CPM per fragment) calculations

Mutant assembled transcripts were aligned to wild-type assembled transcripts via MAFFT v7 [132] in Benchling and regions of shared sequence were identified. Genomic coordinates of these shared exon regions were used to create an annotation file (BED format) for each gene. BAM files (from uniquely aligned RNA-seq reads previously mapped to the W22 genome) for each of the three biological replicates of mutant and W22 control were converted to BED format (BEDTools bamtobed [133]). The RNA-seq mapped read BED file of each mutant or control biological replicate was intersected with the shared exon read annotation file to obtain new read counts (BEDTools intersect-force strandedness [133]). Read counts were then normalized by the effective library size with edgeR v3.28.0 [125] to give CPMs (Counts Per Million reads) for each gene in each sample. Transcript abundance was calculated by averaging the CPMs from all three biological replicates of each mutant allele or W22 control (Table S5). The coordinates of shared sequence regions between mutant and wild-type transcripts usually span multiple exons; therefore, we refer to this transcript abundance calculation as CPM per fragment. Linear regression was used to calculate R^2 correlation values between mutant and wild-type transcript abundance (lm function in R [134]).

Tissue-specific expression patterns of mutant transcripts

Five genes (two independent mutant alleles each) with variable levels of wild-type gene expression across multiple tissues were selected to analyze by RT-qPCR. To examine tissue-specific patterns of expression for each of the 10 mutant alleles relative to wild-

type, 7 different tissues at various stages of maize development (immature to mature) were selected to sample: imbibed embryo, coleoptile tip, seedling shoot, seedling radical root, flag leaf, tassel stem and immature ear spikelet. Three biological replicates of each mutant allele and W22 control were sampled for RNA from each of the seven tissues. The same sampling methods listed above in the Plant Material section were used to sample imbibed embryo, coleoptile tip and tassel stem tissues. Coleoptile tips and radical roots were sampled (9:00) from the same plants, from seeds 6 DAS in 30C dark conditions, and 6 tips and roots were pooled for each biological replicate. Whole shoots from three seedlings 10 DAS in 16 h light 28°C, 8 h dark 24°C growth chamber conditions were sampled (9:00) and pooled for one biological replicate. Along with tassel stems (11:30), unfertilized ear spikelets and flag leaves were sampled (10:30) at anthesis and tissue from three plants was pooled for one biological replicate. Unfertilized ear spikelets were sampled by trimming the terminal 4 cm of the ear and collecting the next 2 cm. For flag leaves, the terminal 15 cm of the flag leaf blade was collected.

Total RNA was extracted from ~100 mg of tissue per sample using TRIzol™ Reagent (Thermo Fisher, Cat # 15596026), DNase treated (TURBO DNA-free™ Kit, Thermo Fisher, Cat # AM1907), and quantified with Quant-iT RiboGreen RNA Assay Kit (Thermo Fisher, Cat # R11490). RNA from imbibed embryos was diluted to 50 ng/ul and RNA from all other tissues was diluted to 75 ng/ul prior to RT-qPCR. Primers for RT-qPCR were designed with specificity to amplify shared gene sequence (~300 bp) between mutant and wild-type transcript assemblies: regions 5' of *Mu* sequence in mutant gene promoter-*Mu* transcripts and regions 3' of *Mu* sequence in mutant *Mu* promoter transcripts (Table S6). The Luna® Universal One-Step RT-qPCR Kit (New England Biolabs, Cat # E3005X) was used to run all RT-qPCR reactions. For each primer-set and tissue, RT-qPCR was performed on three technical replicates of each biological replicate for three biological replicates of the mutant allele and control. Technical replicate Ct values were averaged for each biological replicate. Delta Ct (dCt) values were calculated by the difference between the gene Ct value and the Ct value of the selected maize housekeeping gene, *Ubiquitin Carrier Protein* (*Zm00004b005988*) [135] for each biological replicate. Then, dCt values for all three biological replicates were averaged to give a final dCt value for each transcript.

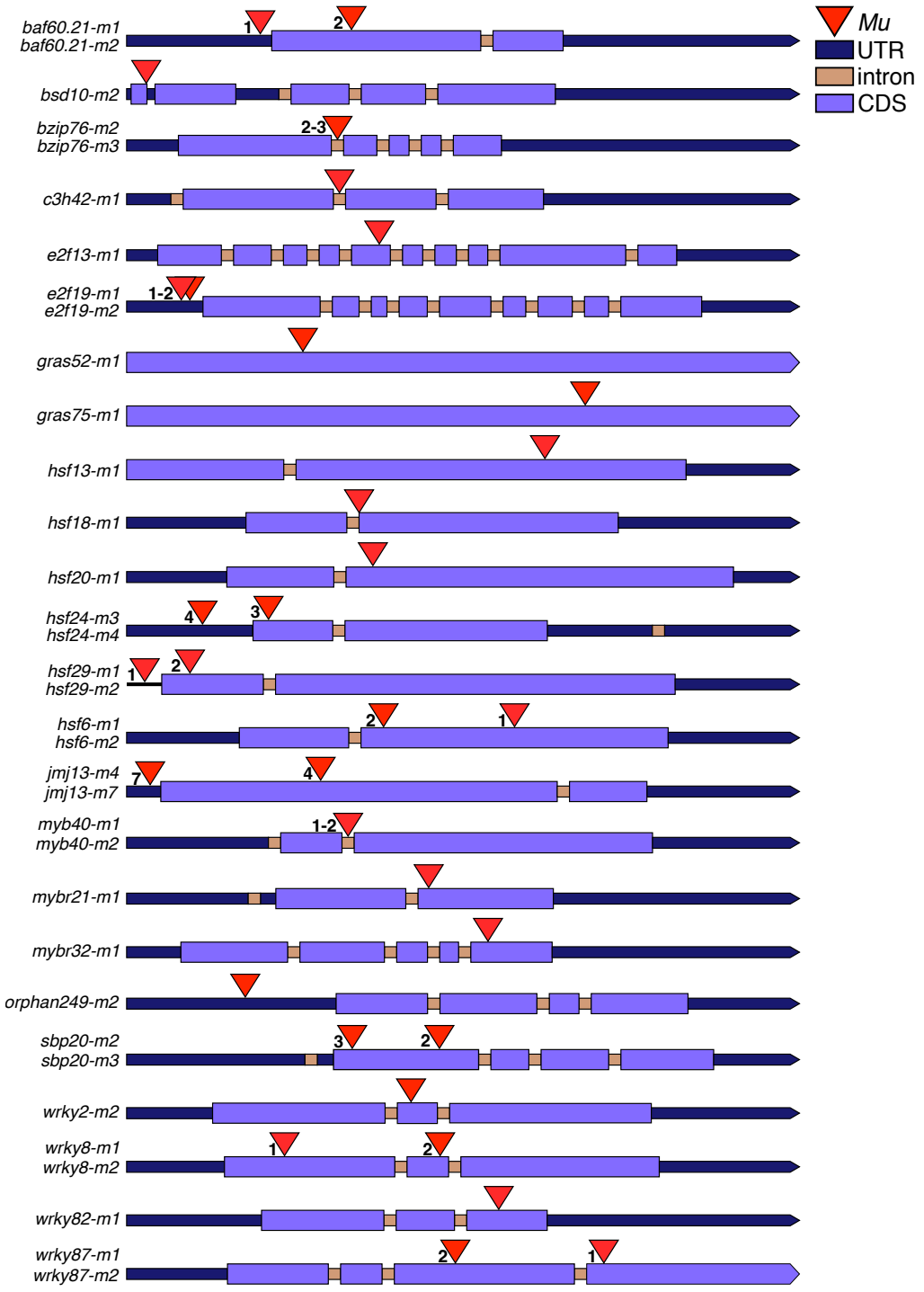


Figure 1. Schematic of *Mutator* insertion locations for 35 UniformMu mutants. The W22 gene models are indicated by different colors/shapes to represent UTRs, coding sequence and introns. The UTRs and CDS' for each gene model are scaled proportionally, but introns are not to scale. *Mu* transposon insertions are indicated by red triangles and independent mutant alleles from the same gene are depicted by numbers. The *BSD10; Zm00004b040474* gene model is based on the B73v4; *Zm00001d026518* gene annotation due to a fused gene annotation in W22.

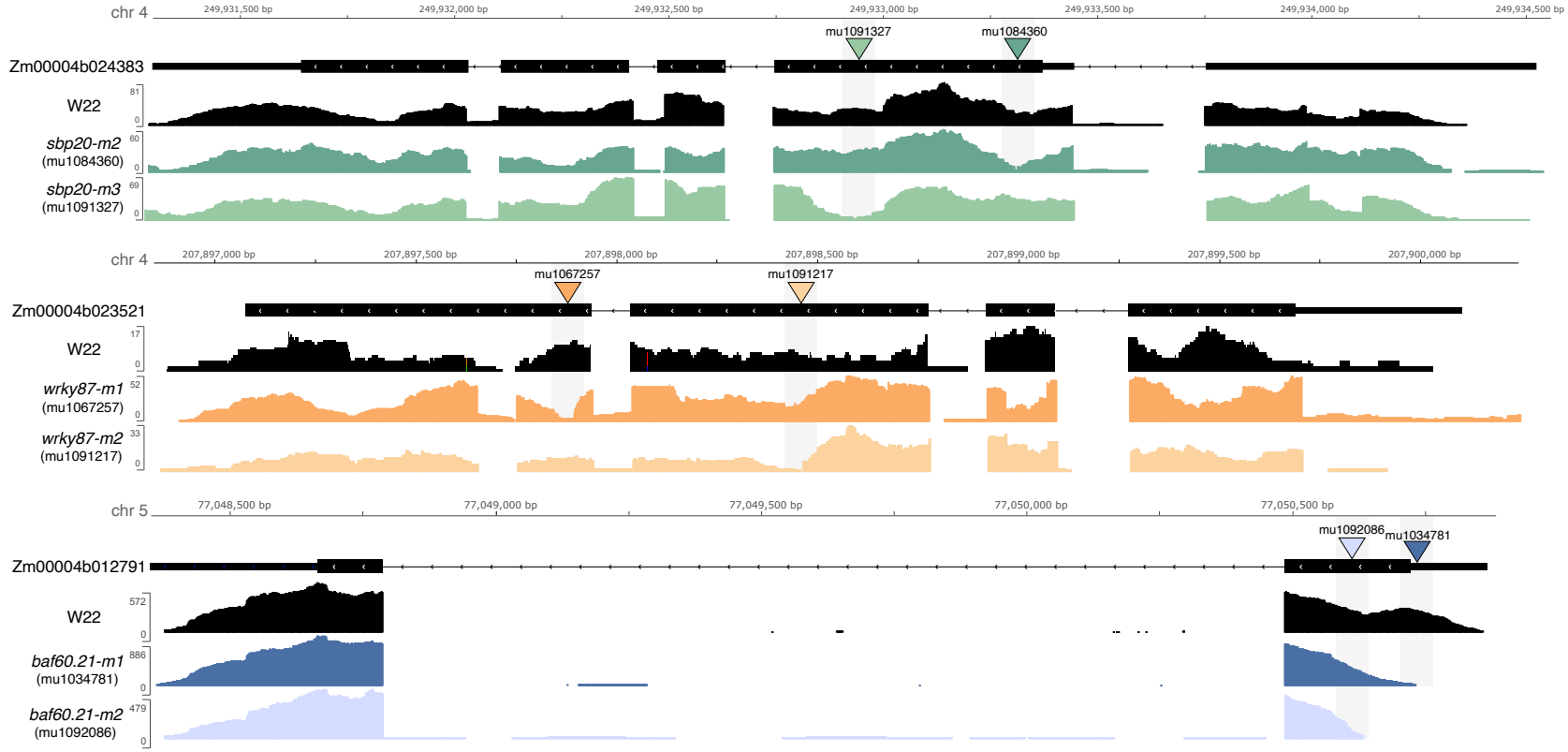


Figure S1. Visualization of wild-type W22 and mutant allele RNA-seq read coverage for three genes. IGV was utilized to visualize the coverage of RNA-seq reads for one biological replicate of each mutant allele for three genes that each have two mutant alleles; *SBP20* (Zm00004b024383), *WRKY87* (Zm00004b023521) and *BAF60.21* (Zm00004b012791). The mutant alleles: *SBP20* (green), *sbp20-m2* (mu1084360) and *sbp20-m3* (mu1091327); *WRKY87* (orange), *wrky87-m1* (mu1067257) and *wrky87-m2* (mu1091217); *BAF60.21* (blue), *baf60.21-m1* (mu1034781) and *baf60.21-m2* (mu1092086) all show reduced coverage in the regions flanking their respective annotated *Mu* insertions (triangles) in the mutant samples compared to W22 wild-type (black).

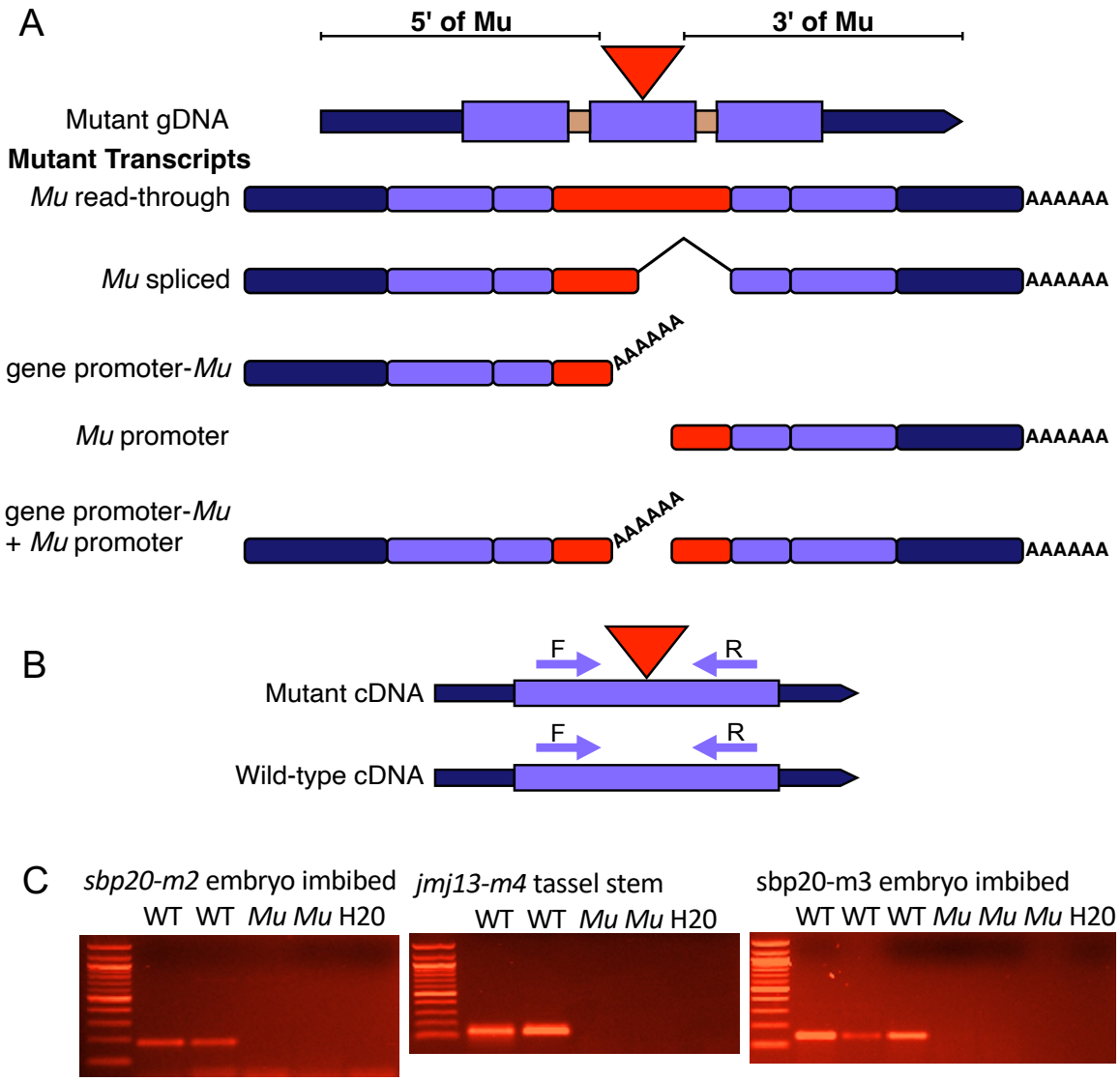


Figure 2. Potential *Mu* insertion allele transcript structures. A) Schematic of 5 mutant transcript structures that could result from a *Mu* transposon insertion. Potential transcripts include: a transcript with sequences 5' and 3' of the *Mu* insertion and all or a portion of *Mu* retained (*Mu* read-through), a transcript with sequence 5' and 3' of the *Mu* insertion and partial *Mu* sequence retained due to alternative splicing (*Mu* spliced), a partial transcript initiating from the gene promoter and terminating in *Mu* (gene promoter-*Mu*), a partial transcript initiating within *Mu* and reading through the 3' gene sequence (Mu initiated), or both the gene promoter-Mu and Mu promoter transcripts. **B)** Schematic of RT-PCR primer design to test for *Mu* read-through and *Mu* spliced transcripts with gene specific primers (F and R) flanking the annotated *Mu* insertion site in mutant and wild-type alleles. A larger PCR product will be amplified in the mutant allele relative to wild-type if all or a portion of *Mu* is retained. No product will be amplified in the mutant if there are transcripts initiating or terminating in *Mu*. **C)** RT-PCR gels of gene-specific primers flanking Mu for at least two biological replicates of 3 mutant and wild-type alleles. All three alleles lack mutant cDNA amplification relative to wild-type.

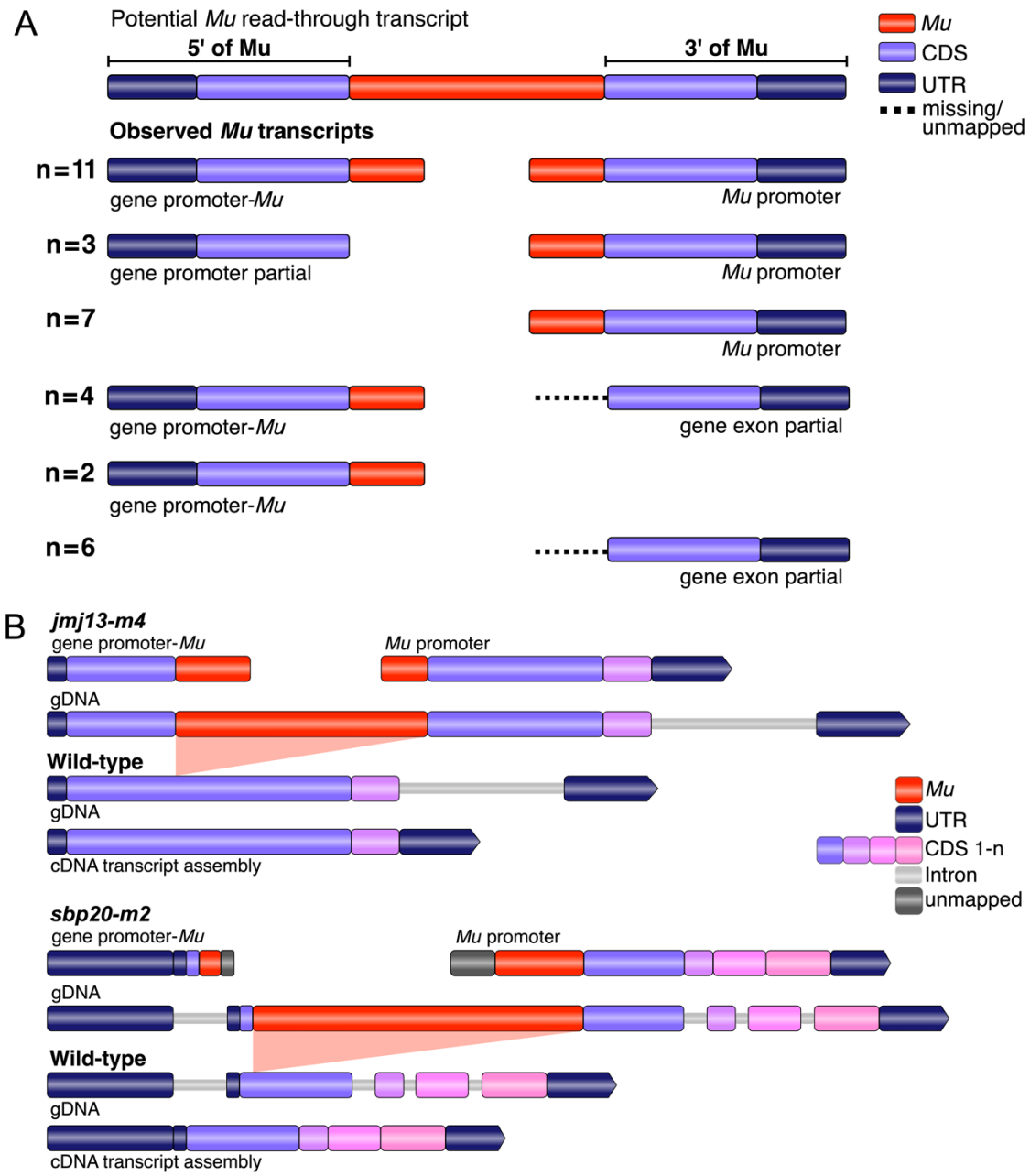


Figure 3. Schematic of *de novo* transcript assemblies for 33 mutant alleles. A) Mutant allele transcripts are referenced relative to the sequences 5' and 3' of the *Mu* insertion. No full-length transcripts with retained *Mu* sequences were identified. The observed transcripts could include separate transcripts representing sequences both 5' and 3' of the Mu insertion or just sequences covering one side of the gene. We also identified some examples of transcripts with partial regions of retain *Mu* sequence. The observations were grouped into 6 types of observations that are illustrated with schematics and the total number of alleles for each type is indicated. **B)** The specific transcripts that are observed for two of the mutant alleles are shown in detail. The wild-type transcript assembly is shown to indicate we recovered the annotated wild-type cDNA with our short-read assembly. Both *jmj13-m4* and *sbp20-m2* have evidence for two transcripts assembled, one transcript initiating from the normal gene promoter with premature termination in *Mu* (5' of *Mu*) and the other initiating within *Mu* and reading through the 3' end of the gene (3' of *Mu*). All *de novo* transcripts were assembled with TRINITY.

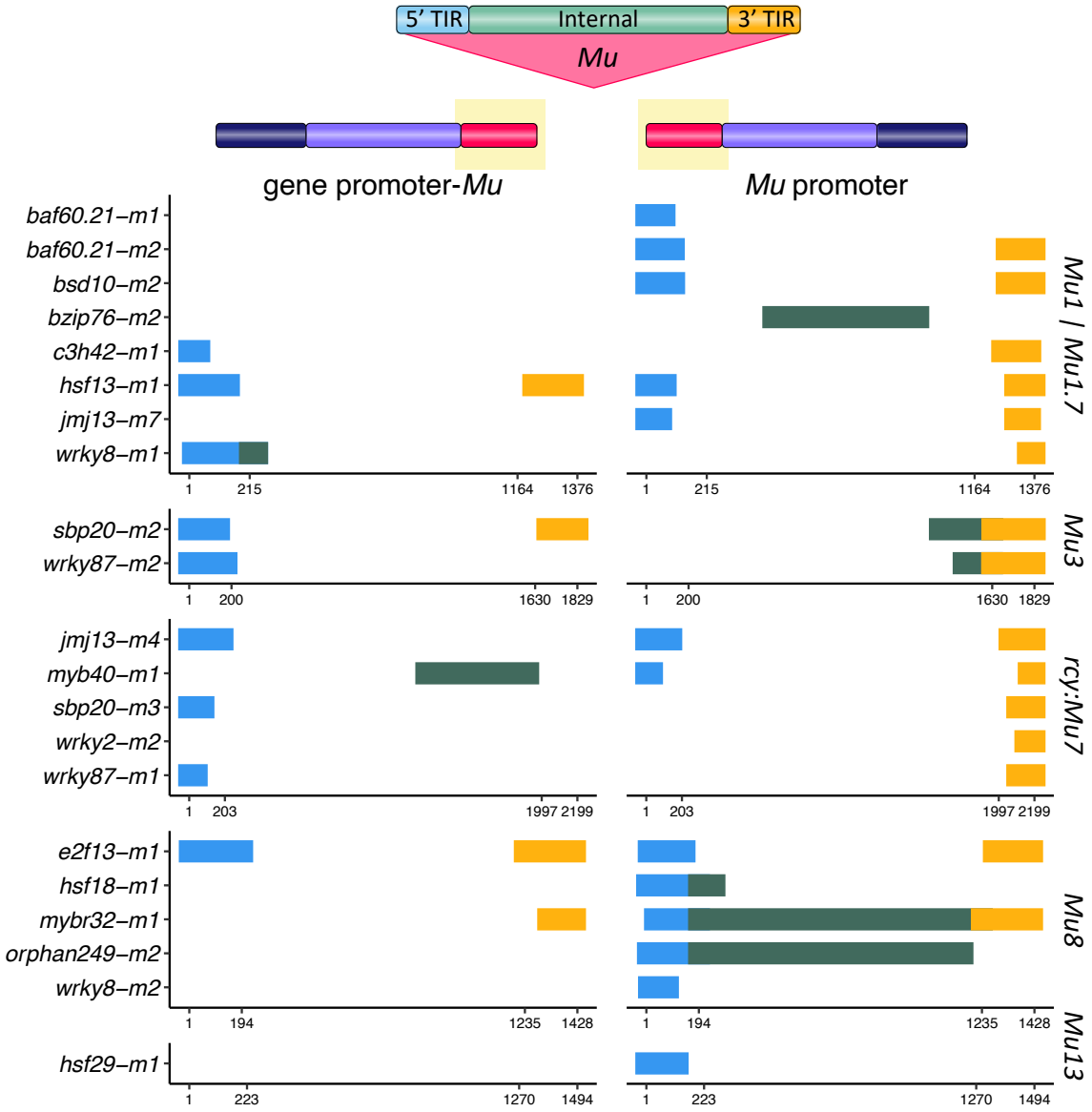
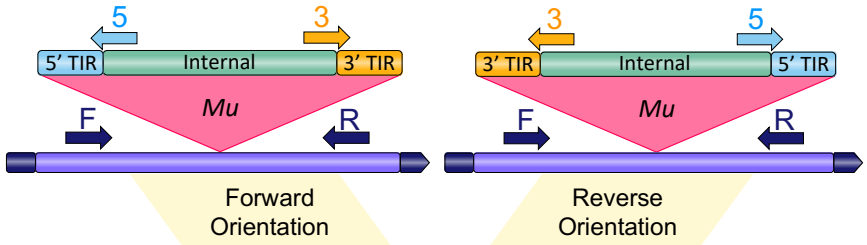


Figure 4. Assessment of *Mu* element identity and orientation. The *Mu* sequence from the de novo assembled transcripts was aligned to representative *Mu* element sequences (indicated to the right of the plots). The left panel shows alignments of the *Mu* sequence from the transcript initiated at the gene promoter while right panel shows alignments of *Mu* sequence from *Mu* initiated transcripts that include the 3' portion of the gene. The alleles are grouped based upon which *Mu* element had the greatest similarity and the plots indicated the position of the aligned sequence within the *Mu* element. *Mu1* and *Mu1.7* are plotted together due to sequence similarity between TIR regions. The *Mu* sequence segments are colored by the alignment position relative to *Mu* element features: 5'TIR (blue), internal sequence (dark green), and 3'TIR (gold).

A



B

Allele	*Mu*	F 5	3 R	F 3	5 R	DIR
sbp20-m2	*Mu3*	+	+	NT	NT	F
wrky87-m2	*Mu3*	+	+	NT	NT	F
sbp20-m3	*Mu7*	+	+	NT	NT	F
jmj13-m4	*Mu7*	+	+	NT	NT	F
wrky87-m1	*Mu7*	+	+	NT	NT	F
wrky8-m1	*Mu1.7*	+	+	NT	NT	F
hsf13-m1	*Mu1.7*	+	+	−	−	F
e2f13-m1	*Mu8*	+	NT	NT	−	F
hsf18-m1	*Mu8*	−	NT	NT	+	R
mybr32-m1	*Mu8*	−	NT	NT	+	R
wrky2-m2	*Mu7*	NT	+	−	NT	F
wrky8-m2	*Mu8*	−	NT	NT	+	R
baf60.21-m1	*Mu1.7*	−	NT	NT	+	R
baf60.21-m2	*Mu1.7*	+	NT	NT	−	F
jmj13-m7	*Mu1.7*	+	NT	NT	−	F
myb40-m1	*Mu7*	NT	+	−	NT	F
orphan249-m2	*Mu8*	+	NT	NT	−	F
c3h42-m1	*Mu1.7*	+	+	−	−	F
bsd10-m2	*Mu1*	NT	−	+	NT	R

C

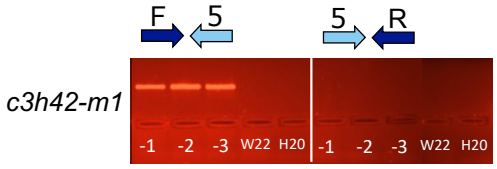


Figure S2. Determination of Mu element identity and orientation by PCR using *Mu* element specific primers. A) Schematic of genomic DNA primer design for identifying the forward or reverse orientation of *Mu*. Gene specific primers (F = Forward, R = Reverse: navy) and predicted *Mu* element primers with specificity to either the 5' TIR (5: light blue) or the 3' TIR (3: yellow) of the predicted *Mu* element were used to complete PCR for each mutant allele. The presence of PCR amplicons from F-5 and R-3 indicate orientation of the *Mu* element in the forward direction (5'TIR to 3'TIR) and the presence of PCR amplicons from F-3 and R-5 indicate orientation of the Mu element in the reverse direction (3'TIR to 5'TIR). **B)** The PCR results of 19 mutant alleles are depicted in a table by amplicon presence (+), absence (-), or not tested (NT). For the alleles with predictions of *Mu* element identity and orientation we only tested the predicted Mu element. For other alleles we tested a variety of primers and both potential orientations and only show the results for the *Mu* element that provided amplification. *bsd10-m2* may be either Mu1 or Mu1.7. **C)** Gel image of PCR amplicons for *c3h42-m1* biological replicates (-1, -2, -3), F-5 presence (+) and R-5 absence (-), and W22 wild-type and water as controls.

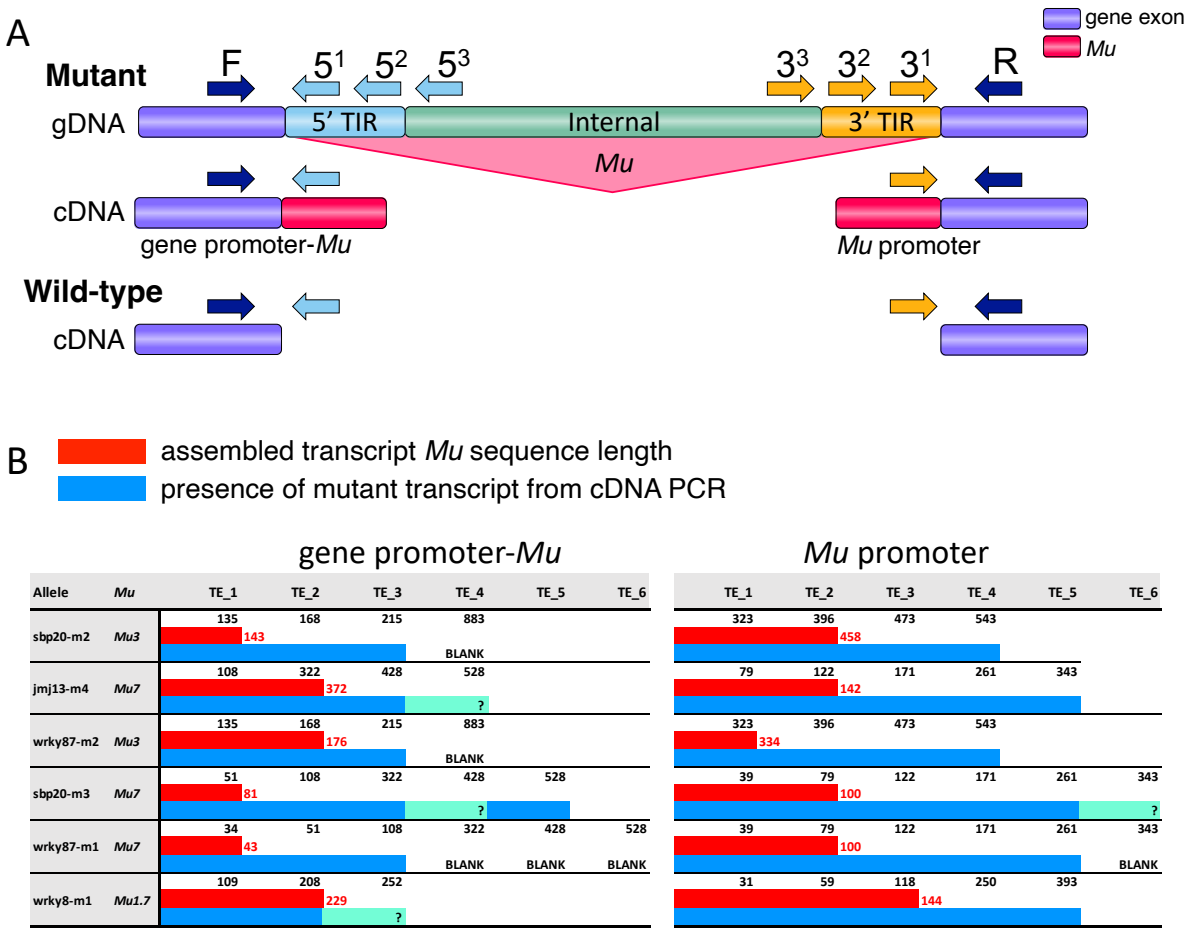


Figure 5. Definition of transcript boundaries using RT-PCR method. A) A schematic of how RT-PCR primers were designed to test the extent of the transcribed *Mu* sequence relative to the predictions of the mutant allele transcript assemblies. Multiple *Mu* primers were designed with specificity to either the 5' or 3' regions of each *Mu* element tested by RT-PCR. For each mutant allele, some *Mu* primers were designed to be located within the transcriptome assembly and some were designed for more internal regions of the *Mu* element. All primer-sets used for RT-PCR were first tested and confirmed to amplify mutant genomic DNA. **B)** A graphic representation of the RT-PCR results is shown. For each mutant allele transcript, the numbers in the table indicate the base pairs into *Mu* where each primer binds. Red highlighting indicates regions expected to amplify based on the transcript assembly (bp of *Mu* sequence listed in Red). Blue highlighting indicates which primers successfully amplified products. TE primers are labeled TE 1-n with primers that are more internal as higher numbers. BLANK indicates that no product was amplified. Teal highlighting with a "?" indicates that presence or absence of amplification is unclear.

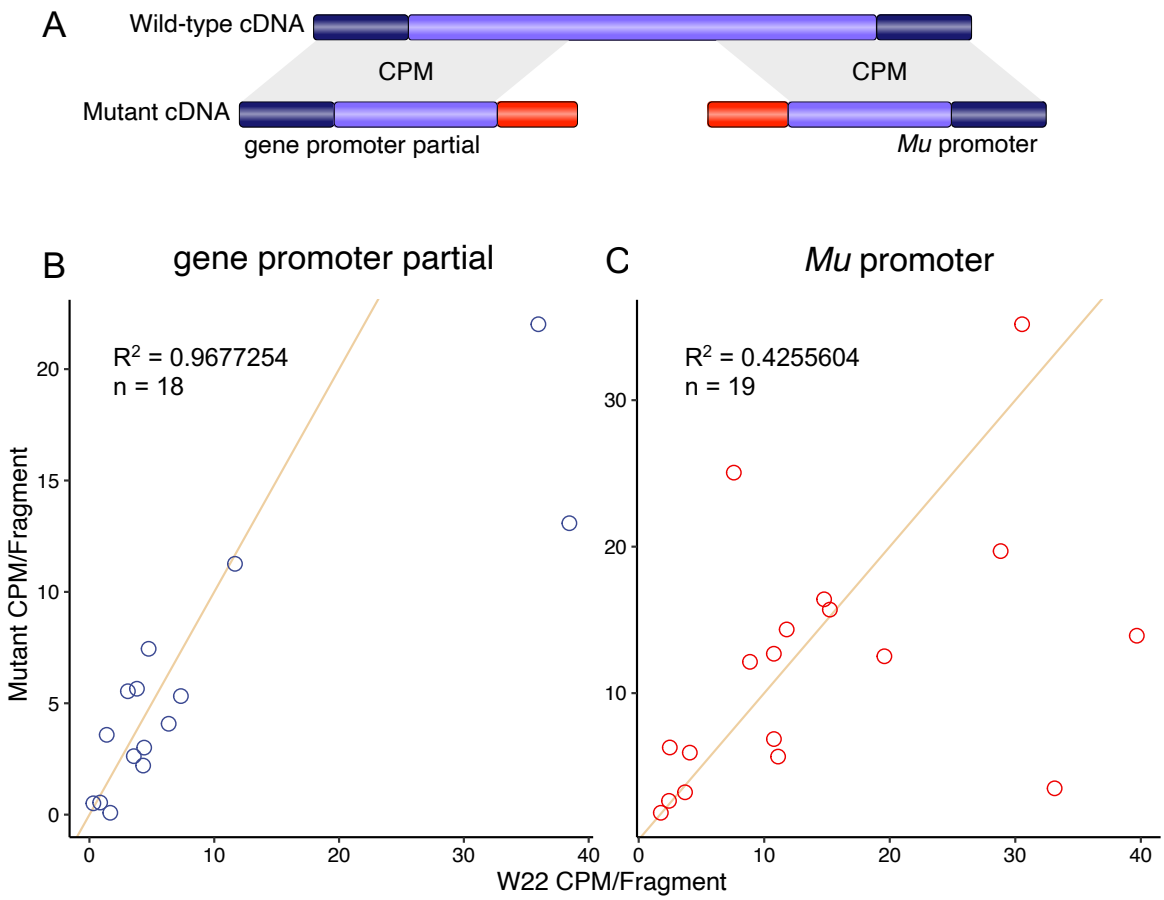


Figure 6. Transcript abundance comparison of mutant and wild-type transcripts. **A)** Schematic of how the expression level of the mutant and wild-type transcripts could be directly compared. Exon regions that are present in both the mutant and wild-type transcript assemblies were identified for each *Mu* allele. Transcript abundance (CPM/fragment) for mutant and wild-type transcripts was calculated from counts of the corresponding RNA-seq sample reads that map to these shared exon regions. A scatter plot of mutant transcript abundance (y-axis coordinates) relative to W22 wild-type transcript abundance (x-axis coordinates) for **B)** gene promoter partial transcripts, including transcripts with termination in *Mu* from 18 mutant alleles (blue circles) and **C)** *Mu* promoter transcripts from 19 mutant alleles (red circles). The lines show the expectation if there was equivalent expression for both alleles (slope = 1). Transcript abundance values for four mutant transcripts (*mybr32-m1*, *gras75-m1* and *wrky82-m1* gene promoter-partial transcripts and *baf60.21-m1 Mu* promoter transcript) are not shown in these plots due to either mutant and/or wild-type abundance > 40 CPM. However, R^2 values were calculated for each transcript abundance comparison from all 18 gene promoter partial transcripts and 19 *Mu* promoter transcripts.

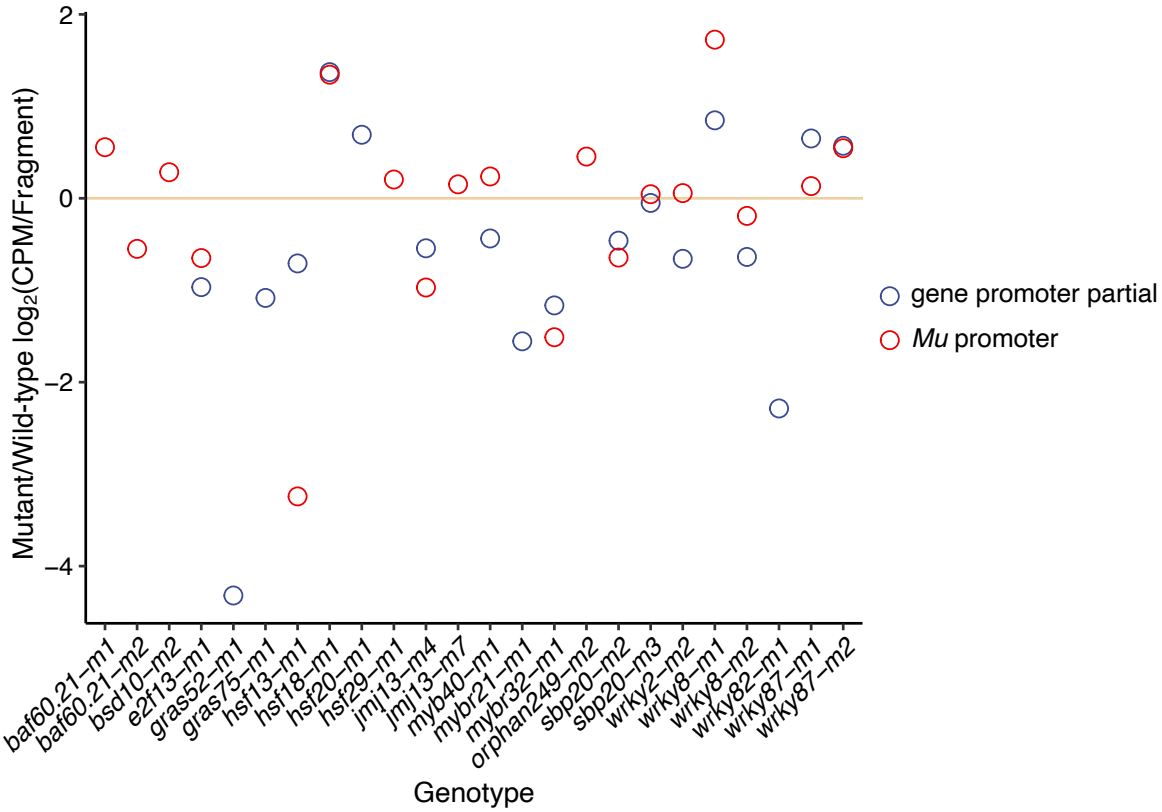


Figure S3. Ratios of mutant to wild-type transcript abundance. A scatter plot of the $\log_2$ normalized ratio of mutant to wild-type transcript abundance (CPM/fragment) is plotted for each mutant allele and the corresponding mutant transcript, gene promoter partial (blue circles) and *Mu* promoter (red circles). Mutant transcripts with expression levels that vary relative to wild-type transcripts are plotted relative to zero with positive values for higher transcript abundance and negative values for lower transcript abundance

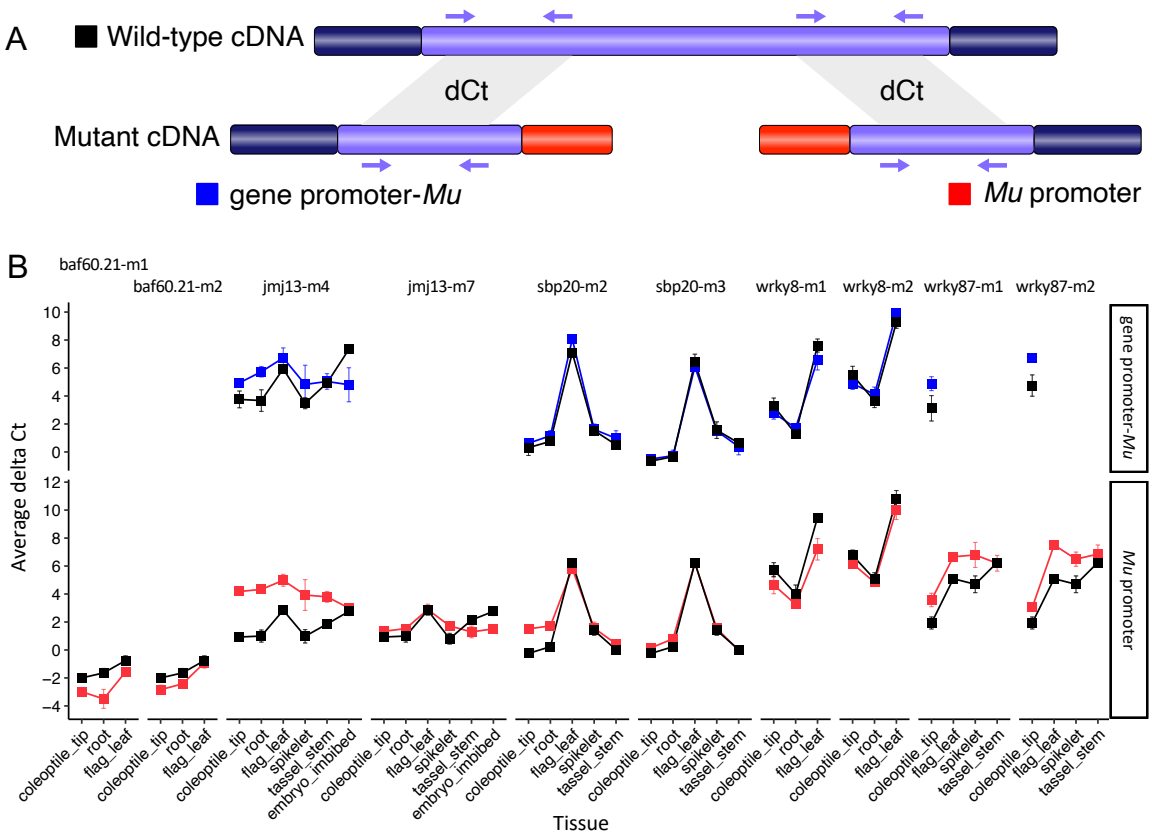


Figure 7. Tissue-specific patterns of expression for mutant transcripts relative wild-type transcripts from RT-qPCR data. A) Schematic of RT-qPCR primer design to test relative patterns of gene expression from mutant and wild-type alleles. Primers with specificity to regions of shared sequence between assembled wild-type and mutant transcripts were used to quantify tissue-specific expression patterns by RT-qPCR. 10 mutant alleles isolated from 5 genes were selected to sample for this analysis. Selecting genes with two independent mutant alleles each provided more power to confirm relative wild-type gene expression patterns across tissues and we focused on genes with variable expression levels in the tissues sampled. **B)** Average delta Ct (dCt) values of mutant gene promoter-*Mu* transcripts (blue) and mutant *Mu* promoter transcripts (red) relative to the values of their corresponding wild-type transcripts (black) are plotted as points each tissue sampled (x-axis tissue order is the same for all alleles). Each data point is the average dCt value of three biological replicates by three technical replicates. Trendlines across tissues are plotted to display expression patterns. dCt values are inversely proportional to relative transcript abundance and greater dCt values indicate lower relative transcript abundance.

Table S1. Genomic data for 35 *Mu*-insertion mutant alleles isolated from the UniformMu population in maize.

			UniformMu Mutant Insertion Maize Genomic Data			
Name	**B73v4_gid**	**W22_gid**	**Mutant Allele**	**W22_MuID**	**B73v4_MuID**	**Stock**
BAF60.21	Zm00001d015127	Zm00004b012791	baf60.21-m1	chr5_77050725	mu1034781	UFMu-03236
BAF60.21	Zm00001d015127	Zm00004b012791	baf60.21-m2	chr5_77050622	mu1092086	UFMu-11153
BSD10	Zm00001d026518	Zm00004b040474	bsd10-m2	chr10_145400162	mu1072546	UFMu-03453
BZIP76	Zm00001d036736	Zm00004b029476	bzip76-m2	chr6_95884230	mu1083967	UFMu-10010
BZIP76	Zm00001d036736	Zm00004b029476	bzip76-m3	chr6_95884945	mu1081677	UFMu-09463
C3H42	Zm00001d008356	Zm00004b024707	c3h42-m1	chr8_5818897	mu1068676	UFMu-08608
E2F19	Zm00001d027709	Zm00004b000391	e2f19-m1	chr1_12237717	mu1040458	UFMu-04300
E2F19	Zm00001d027709	Zm00004b000391	e2f19-m2	chr1_12237755	mu1080409	UFMu-09504
HSF24	Zm00001d032923	Zm00004b004268	hsf24-m3	chr1_245656248	mu1065831	UFMu-08727
HSF24	Zm00001d032923	Zm00004b004268	hsf24-m4	chr1_245656475	mu1037205	UFMu-03655
JMJ13	Zm00001d052933	Zm00004b023587	jmj13-m4	chr4_210612840	mu1081210	UFMu-09466
JMJ13	Zm00001d052933	Zm00004b023587	jmj13-m7	chr4_210613404	mu1041192	UFMu-04767
MYB40	Zm00001d040621	Zm00004b016719	myb40-m1	chr3_55921183	mu1090929	UFMu-11189
MYB40	Zm00001d040621	Zm00004b016719	myb40-m2	chr3_55920569	mu1043293	UFMu-04991
E2F13	Zm00001d052288	Zm00004b023063	e2f13-m1	chr4_192230417	mu1086360	UFMu-11121
GRAS52	Zm00001d002573	Zm00004b006535	gras52-m1	chr2_15799332	mu1037818	UFMu-03743
GRAS75	Zm00001d006701	Zm00004b009784	gras75-m1	chr2_213696662	mu1080849	UFMu-09435
HSF13	Zm00001d027757	Zm00004b000433	hsf13-m1	chr1_13648451	mu1085337	UFMu-10587
HSF18	Zm00001d016255	Zm00004b013608	hsf18-m1	chr5_153354999	mu1086526	UFMu-10749
HSF20	Zm00001d026094	Zm00004b040094	hsf20-m1	chr10_136927897	mu1085720	UFMu-10752
HSF29	Zm00001d016520	Zm00004b013825	hsf29-m1	chr5_165555661	mu1023451	UFMu-02314
HSF29	Zm00001d016520	Zm00004b013825	hsf29-m2	chr5_165555817	mu1083642	UFMu-10398
HSF6	Zm00001d016674	Zm00004b013941	hsf6-m1	chr5_171668444	mu1048425	UFMu-06347
HSF6	Zm00001d016674	Zm00004b013941	hsf6-m2	chr5_171667932	mu1056797	UFMu-07611
MYBR21	Zm00001d008602	Zm00004b024904	mybr21-m1	chr8_14018025	mu1042351	UFMu-04838
MYBR32	Zm00001d029963	Zm00004b002134	mybr32-m1	chr1_96615987	mu1076168	UFMu-09083
ORPHAN249	Zm00001d038846	Zm00004b031116	orphan249-m2	chr6_163275141	mu1029949	UFMu-02921
SBP20	Zm00001d053890	Zm00004b024383	sbp20-m2	chr4_249933323	mu1086430	UFMu-10892
SBP20	Zm00001d053890	Zm00004b024383	sbp20-m3	chr4_249932948	mu1091327	UFMu-11256
WRKY2	Zm00001d016052	Zm00004b013450	wrky2-m2	chr5_141290461	mu1025953	UFMu-11813
WRKY8	Zm00001d053369	Zm00004b023980	wrky8-m1	chr4_235052564	mu1048159	UFMu-06456
WRKY8	Zm00001d053369	Zm00004b023980	wrky8-m2	chr4_235053087	mu1077370	UFMu-08953
WRKY82	Zm00001d038843	Zm00004b031112	wrky82-m1	chr6_163095668	mu1081611	UFMu-09469
WRKY87	Zm00001d052847	Zm00004b023521	wrky87-m1	chr4_207897892	mu1067257	UFMu-08437
WRKY87	Zm00001d052847	Zm00004b023521	wrky87-m2	chr4_207898459	mu1091217	UFMu-12044

Table S2. Gene expression values for 24 transcription factor genes (B73v4) in different tissues (Zhou et al. 2019). The expression value (CPM) for each of the 22 TFs was assessed based on prior sampling of tissues or developmental stages in B73 (Zhou et al. 2019). Values highlighted indicate the predicted expression level of TF genes in tissues sampled for RNA-seq in this study.

		24 h	6 DAS		seedling 11 DAS			v12					v14	0 DAP							14 DAP		27 DAP	
TF	RNAseq Tissue	embryo imbibed	coleoptile tip	radicle root	leaf	root	meristem	blade leaf	sheath	internode	tassel	auricle	ear	silk	spikelet	husk	tassel stem	floret	flag leaf	root	kernel	endosperm	embryo	endosperm
WRKY2	coleoptile_tip	3.0	11.9	6.2	1.4	3.7	6.2	0.6	0.9	1.7	2.5	1.2	1.5	1.4	2.1	1.3	2.6	0.9	1.2	0.3	3.1	2.4	3.9	1.3
MYB40	coleoptile_tip	4.9	45.7	21.2	7.9	1.6	19.1	5.7	4.1	0.8	14.1	12.8	10.4	6.8	9.8	4.0	1.3	8.5	59.8	2.0	22.3	23.9	5.5	26.2
E2F13	coleoptile_tip	0.1	20.7	9.6	0.0	6.1	34.2	0.0	0.0	0.0	5.9	0.0	34.8	0.1	2.9	0.1	0.1	0.1	0.1	2.3	1.1	0.5	6.5	0.0
GRAS52	embryo_imbibed	114.8	5.2	5.4	23.7	13.7	10.5	9.0	17.9	16.2	12.3	20.5	9.7	3.9	12.9	13.8	28.8	12.4	23.6	10.5	8.6	11.8	5.0	48.5
GRAS75	embryo_imbibed	163.0	32.4	21.9	17.7	7.3	14.8	32.9	27.0	9.6	7.3	18.2	18.3	12.5	16.4	16.1	61.0	19.8	71.3	4.4	47.6	62.1	20.8	58.2
MYBR21	embryo_imbibed	183.3	16.9	28.7	66.4	101.0	20.6	28.1	42.5	55.7	15.2	31.5	23.5	29.4	27.3	31.0	71.9	35.7	40.9	26.3	21.1	20.6	4.8	29.7
HSF18	embryo_imbibed	74.2	10.2	2.5	22.0	34.7	7.0	0.8	5.4	6.1	0.2	3.8	0.1	0.7	0.3	0.2	0.4	0.5	1.3	63.7	2.3	4.4	8.4	7.3
HSF29	embryo_imbibed	15.1	15.2	15.7	15.1	20.9	11.3	23.2	26.5	40.2	22.3	25.8	13.2	30.2	24.5	32.1	29.1	17.7	13.2	21.1	14.5	8.2	14.1	9.8
HSF6	embryo_imbibed	175.7	0.8	5.3	1.7	1.1	0.5	32.4	21.3	10.3	5.3	11.7	4.7	11.0	6.2	7.9	16.8	7.8	45.0	12.0	3.9	2.4	16.1	3.0
HSF20	embryo_imbibed	14.0	1.1	0.4	15.7	7.7	9.7	28.2	11.9	9.6	9.0	13.6	4.7	4.8	5.3	4.7	4.0	5.6	23.4	2.4	4.3	12.4	3.0	10.0
WRKY82	embryo_imbibed	43.5	7.0	15.8	31.0	16.6	1.5	7.7	12.3	18.3	1.5	11.9	0.5	1.4	2.7	8.9	14.3	5.3	10.8	43.6	4.1	0.1	0.2	0.6
WRKY87	embryo_imbibed	40.1	0.4	0.1	12.0	0.5	0.3	9.5	2.0	0.4	0.2	1.4	0.3	0.0	0.1	1.6	1.1	1.6	14.1	0.2	0.3	0.1	0.6	0.1
WRKY8	embryo_imbibed	6.2	3.9	3.1	0.9	4.7	4.6	0.9	2.3	2.6	1.9	3.4	2.6	2.6	1.7	1.6	3.1	1.2	2.7	1.1	1.3	1.0	4.7	0.6
SBP20	embryo_imbibed	37.5	11.8	13.5	5.3	11.0	23.5	1.4	10.4	18.7	16.9	7.5	11.8	2.2	8.5	4.8	17.0	7.9	1.7	37.3	3.7	2.3	6.1	0.8
BZIP76	embryo_imbibed	10.8	2.7	5.2	13.0	2.6	2.0	2.3	1.5	6.7	2.9	0.6	3.1	13.8	5.0	1.4	7.0	4.3	5.3	5.8	2.9	2.2	1.4	1.6
BSD10	embryo_imbibed	19.0	8.3	8.7	8.3	10.2	8.5	12.7	13.5	8.8	11.4	11.8	13.9	12.6	15.0	10.8	8.7	8.4	6.2	10.9	11.3	14.5	15.4	17.2
ORPHAN249	embryo_imbibed	57.0	7.5	26.6	34.5	30.0	7.8	28.6	38.1	24.7	14.3	30.8	20.4	28.2	27.0	14.7	31.3	22.9	61.5	38.9	8.5	2.8	4.4	1.9
E2F19	seedling_leaf	15.8	9.7	10.3	17.0	9.9	15.8	13.5	13.5	11.7	12.6	11.0	23.1	14.6	14.6	10.0	15.2	8.8	22.0	11.4	14.7	14.1	13.0	19.9
HSF13	seedling_leaf	12.4	2.6	5.2	454.1	11.4	1.7	418.9	374.1	9.7	1.1	111.3	1.6	21.2	1.2	99.3	73.6	46.1	650.7	11.2	1.0	0.0	0.2	0.3
MYBR32	seedling_leaf	25.4	22.4	40.5	299.6	136.9	29.4	35.4	158.2	486.2	25.6	189.4	71.8	208.4	55.4	199.4	233.8	114.5	89.7	263.6	20.6	3.1	37.4	5.3
HSF24	tassel	61.0	32.3	39.4	76.4	52.7	120.3	457.3	250.3	292.2	96.7	188.4	80.7	82.7	100.1	73.7	109.6	89.0	400.1	155.9	167.1	208.2	265.2	121.1
C3H42	tassel_stem	3.9	7.4	10.4	2.2	4.6	9.5	3.5	5.2	27.3	7.5	3.9	6.9	8.2	7.2	6.7	9.6	2.8	4.7	7.8	9.1	11.2	4.9	6.5
BAF60.21	tassel_stem	14.6	22.8	20.9	4.2	12.8	40.2	14.1	14.7	8.2	11.7	14.1	19.8	14.1	10.8	13.3	16.3	11.3	24.1	14.9	7.5	6.1	16.7	3.9
JMJ13	tassel_stem	27.5	19.5	14.9	0.9	11.5	21.1	6.1	8.2	6.9	7.0	7.0	18.3	10.7	11.4	6.6	6.8	6.5	7.8	8.1	20.1	23.5	13.9	16.0

Table S3. Mutant allele transcriptome data obtained, transcript abundance, and transcriptome assembly predicted transcript structure. RNA-seq CPM and FPKM for mutant and wild-type W22, DE data averaged across biological replicates (N), log2fc: log2 fold change of mutant to control, lfcSE: log fold change standard error and FDR adjusted p-value.

				Trancript Abundance										
	Transcriptome Data Obtained			mutant			W22			Gene DE Data				
Genotype	Pedigree	Tissue	Illumina	FPKM	CPM	N	FPKM	CPM	N	log2fc	lfcSE	p-value	DE	Transcript Structure- Figure 3A
baf60.21-m1	BC2S2	tassel_stem	PE 150bp	45.60	37.49	3.00	32.79	26.96	3	0.480	0.114	8.22E-04	Not_DE	Mu promoter
baf60.21-m2	BC2S2	tassel_stem	PE 150bp	18.52	15.22	3.00	32.79	26.96	3	-0.829	0.118	1.16E-09	Not_DE	Mu promoter
bsd10-m2	S2	embryo_imbibed	PE 150bp	12.62	22.79	3.00	16.19	29.24	4	-0.438	0.108	4.86E-04	Not_DE	Mu promoter
bzip76-m2	BC2S2	leaf	PE 150bp	4.31	7.92	3.00	4.04	7.41	4	0.064	0.217	0.90064	Not_DE	Mu promoter
bzip76-m3	BC2S2	leaf	PE 150bp	2.07	3.80	3.00	4.04	7.41	4	-0.954	0.237	0.0011	Not_DE	gene exon partial
c3h42-m1	BC2S2	tassel_stem	PE 150bp	7.20	5.05	3.00	5.31	3.73	3	0.443	0.173	0.06283	Not_DE	gene promoter-Mu | Mu promoter
e2f13-m1	BC2S2	coleoptile_tip	PE 150bp	3.15	5.77	3.00	4.86	8.90	4	-0.607	0.159	0.00362	Not_DE	gene promoter-Mu | Mu promoter
e2f19-m1	BC2S2	leaf	PE 150bp	2.43	4.13	3.00	5.43	9.22	4	-1.125	0.217	8.47E-06	DE_Down	gene exon partial
e2f19-m2	BC2S2	leaf	PE 150bp	4.01	6.82	3.00	5.43	9.22	4	-0.471	0.203	0.11291	Not_DE	gene exon partial
gras52-m1	BC2S2	embryo_imbibed	PE 150bp	0.38	0.72	2.00	6.59	12.33	4	-4.072	0.355	4.83E-28	DE_Down	gene promoter-Mu | gene exon partial
gras75-m1	BC2S2	embryo_imbibed	PE 150bp	32.49	54.10	3.00	78.36	130.48	4	-1.266	0.142	4.06E-17	DE_Down	gene promoter-Mu
hsf13-m1	BC2S2	leaf	PE 150bp	31.24	45.52	3.00	63.51	92.53	4	-0.989	0.110	4.03E-17	Not_DE	gene promoter-Mu | Mu promoter
hsf18-m1	BC2S2	embryo_imbibed	PE 150bp	5.40	7.69	3.00	1.98	2.83	4	1.373	0.290	2.90E-05	DE_Up	gene promoter partial | Mu promoter
hsf20-m1	S3	embryo_imbibed	PE 150bp	0.71	0.89	3.00	0.49	0.61	4	0.514	0.692	0.63301	Not_DE	gene promoter-Mu | gene exon partial
hsf24-m3	S3	tassel	SE 50bp	21.03	51.22	3.00	16.89	41.15	3	0.045	0.168	0.87331	Not_DE	gene promoter-Mu | gene exon partial
hsf24-m4	S2	tassel	SE 50bp	8.72	21.24	3.00	16.89	41.15	3	-1.104	0.172	9.37E-09	DE_Down	gene exon partial
hsf29-m1	S3	embryo_imbibed	PE 150bp	17.78	32.75	3.00	15.11	27.83	4	0.201	0.168	0.44145	Not_DE	Mu promoter
hsf29-m2	S3	embryo_imbibed	PE 150bp	15.17	27.94	3.00	15.11	27.83	4	-0.025	0.169	0.93601	Not_DE	gene exon partial
hsf6-m1	S3	embryo_imbibed	PE 150bp	0.02	0.04	3.00	0.20	0.48	4	-3.748	1.245	0.01579	DE_Down	NA
hsf6-m2	S3	embryo_imbibed	PE 150bp	0.14	0.34	3.00	0.20	0.48	4	-0.479	0.772	0.72818	Not_DE	NA
jmj13-m4	S3	tassel_stem	PE 150bp	4.24	8.90	3.00	6.73	14.14	3	-0.669	0.133	1.13E-04	Not_DE	gene promoter-Mu | Mu promoter
jmj13-m7	BC2S2	tassel_stem	PE 150bp	7.54	15.85	3.00	6.73	14.14	3	0.162	0.128	0.60044	Not_DE	Mu promoter
myb40-m1	S3	coleoptile_tip	PE 150bp	7.95	13.17	3.00	6.98	11.57	4	0.184	0.147	0.89794	Not_DE	gene promoter-Mu | Mu promoter
myb40-m2	S2	coleoptile_tip	PE 150bp	5.29	8.76	3.00	6.98	11.57	4	-0.372	0.151	0.34221	Not_DE	gene exon partial
mybr21-m1	S3	embryo_imbibed	PE 150bp	4.56	14.29	3.00	19.42	60.91	4	-2.075	0.129	1.64E-55	DE_Down	gene promoter-Mu | gene exon partial
mybr32-m1	S2	leaf	PE 150bp	37.42	61.19	3.00	86.66	141.69	4	-1.187	0.083	2.06E-43	DE_Down	gene promoter-Mu | Mu promoter
orphan249-m2	BC1S2	embryo_imbibed	PE 150bp	7.71	17.17	3.00	6.23	13.86	4	0.298	0.239	0.41145	Not_DE	Mu promoter
sbp20-m2	S3	embryo_imbibed	PE 150bp	3.55	9.39	3.00	5.31	14.03	4	-0.620	0.160	8.08E-04	Not_DE	gene promoter-Mu | Mu promoter
sbp20-m3	S2	embryo_imbibed	PE 150bp	5.63	14.90	3.00	5.31	14.03	4	0.065	0.160	0.82448	Not_DE	gene promoter-Mu | Mu promoter
wrky2-m2	S3	coleoptile_tip	PE 150bp	1.23	2.21	3.00	1.28	2.29	4	-0.059	0.258	0.98609	Not_DE	gene promoter partial | Mu promoter
wrky8-m1	S3	embryo_imbibed	PE 150bp	10.99	19.28	3.00	4.08	7.15	4	1.452	0.295	2.40E-05	DE_Up	gene promoter-Mu | Mu promoter
wrky8-m2	S3	embryo_imbibed	PE 150bp	3.07	5.38	3.00	4.08	7.15	4	-0.330	0.301	0.51222	Not_DE	gene promoter partial | Mu promoter
wrky82-m1	S3	embryo_imbibed	PE 150bp	4.40	7.19	3.00	21.01	34.36	4	-2.297	0.262	8.04E-17	DE_Down	gene promoter-Mu
wrky87-m1	S3	embryo_imbibed	PE 150bp	3.39	8.83	3.00	2.30	5.99	4	0.535	0.252	0.10716	Not_DE	gene promoter-Mu | Mu promoter
wrky87-m2	S3	embryo_imbibed	PE 150bp	3.40	8.84	3.00	2.30	5.99	4	0.571	0.252	0.07387	Not_DE	gene promoter-Mu | Mu promoter

Table S4. Mutant allele *Mu* element identity and orientation by gDNA PCR. Table follows the format of Figure S2 with primer sets not tested grayed out, and tested primer sets resulting in amplification—blue, no amplification—pink.

		Forward				Reverse				
		F : Mu 5'		Mu 3' : R		F : Mu 3'		Mu 5' : R		
Allele	***Mu***	**F**	**Mu 5'**	**Mu 3'**	**R**	**F**	**Mu 3'**	**Mu 5'**	**R**	**Orientation**
sbp20-m2	*Mu3*	gsp01	qrp28	qrp21	gsp02					**F**
wrky87-m2	*Mu3*	gsp03	qrp28	qrp21	gsp04					**F**
sbp20-m3	*Mu7*	qrp11	qrp36	qrp62	qrp12					**F**
jmj13-m4	*Mu7*	gsp08	qrp36	qrp62	gsp07					**F**
wrky87-m1	*Mu7*	qrp05	qrp36	qrp62	gsp06					**F**
wrky8-m1	*Mu1.7*	gsp09	qrp41	qrp43	gsp10					**F**
hsf13-m1	*Mu1.7*	qrp41	gsp11	mtp02	gsp12	qrp43	gsp11	qrp41	gsp12	**F**
e2f13-m1	*Mu8*	ump220	mtp10					ump221	mtp10	**F**
hsf18-m1	*Mu8*	ump139	mtp10					ump056	mtp10	**R**
mybr32-m1	*Mu8*	ump042	mtp10					ump125	mtp10	**R**
wrky2-m2	*Mu7*			ump065	qrp63	ump147	qrp63			**F**
wrky8-m2	*Mu8*	ump080	mtp10					ump163	mtp10	**R**
baf60.21-m1	*Mu1.7*	BAF60.21-F3	qrp41					qrp41	BAF60.21-R3	**R**
baf60.21-m2	*Mu1.7*	BAF60.21-F3	qrp41					qrp41	BAF60.21-R3	**F**
jmj13-m7	*Mu1.7*	JMJ13-F9	qrp41					qrp41	JMJ13-m7R9	**F**
myb40-m1	*Mu7*			qrp62	MYB40-R3	MYB40-F1	qrp62			**F**
orphan249-m2	*Mu8*	Orphan249-R9	mtp10					mtp10	Orphan249-F9	**F**
c3h42-m1	*Mu1.7*	C3H42-R1	qrp41	mtp01	C3H42-F1	C3H42-R1	mtp01	qrp41	C3H42-F1	**F**
bsd10-m2	*Mu1*			mtp01	BSD8-R1	BSD8-F1	mtp01			**R**

Table S5. Mutant allele transcript boundaries by RT-PCR. Table follows format of Figure 5B with primer *Mu*-specific primer names listed above the *Mu* sequence amplified and the gene-specific primer used for each allele in the corresponding row. Some *Mu*-specific primers have specificity to both Mu TIRs and can be used with both gene-specific primers to determine transcript boundaries.

gene promoter-*Mu* transcripts

Allele	*Mu*	gene primer	Mu-specific primers TE_1	TE_2	TE_3	TE_4	TE_5	TE_6
sbp20-m2	*Mu3*	gsp01	qrp25 135	qrp27-rc 168	qrp28 215	qrp61 883		
			143					
						BLANK		
jmj13-m4	*Mu7*	gsp08	qrp34 108	qrp64 322	qrp65 428	qrp66 528		
			372					
						?		
wrky87-m2	*Mu3*	gsp03	qrp25 135	qrp27-rc 168	qrp28 215	qrp61 883		
			176					
						BLANK		
sbp20-m3	*Mu7*	qrp11	qrp33 51	qrp34 108	qrp64 322	qrp65 428	qrp66 528	
			81					
						?		
wrky87-m1	*Mu7*	gsp05	qrp37 34	qrp33 51	qrp34 108	qrp64 322	qrp65 428	qrp66 528
			43					
						BLANK	BLANK	BLANK
wrky8-m1	*Mu1.7*	gsp09	qrp44.2 109	qrp40 208	qrp41 252			
			229					
					?			

***Mu* promoter transcripts**

gene primer	Mu-specific primers TE_1	TE_2	TE_3	TE_4	TE_5	TE_6
gsp02	qrp21 323	qrp22 396	qrp23 473	qrp24 543		
	458					
gsp07	qrp30 79	qrp31 122	qrp32 171	qrp62 261	qrp63 343	
	142					
gsp04	qrp21 323	qrp22 396	qrp23 473	qrp24 543		
	334					
qrp12	qrp29 39	qrp30 79	qrp31 122	qrp32 171	qrp62 261	qrp63 343
	100					
						?
gsp06	qrp29 39	qrp30 79	qrp31 122	qrp32 171	qrp62 261	qrp63 343
	100					
						BLANK
gsp10	qrp43 31	qrp42 59	qrp44.2 118	qrp40 250	mtp01 393	
	144					

CHAPTER III. ABSTRACT

Transcription factors (TFs) play important roles in regulation of gene expression and phenotype. A variety of approaches have been utilized to develop gene-regulatory networks (GRNs) to predict the regulatory targets for each TF, such as yeast-one-hybrid (Y1H) screens and gene co-expression network (GCN) analysis. Here we identified potential TF targets and used a reverse genetics approach to test the predictions of several GRNs in maize. Loss-of-function mutant alleles were isolated for 22 maize TFs. These mutants did not exhibit obvious morphological phenotypes. However, transcriptome profiling revealed differentially expressed genes in each of the mutant genotypes. An analysis of expression levels for predicted target genes based on yeast one-hybrid screens identified a small subset of predicted targets that exhibit altered expression levels. The analysis of predicted targets from GCN-based methods found significant enrichments for prediction sets of some TFs, but the majority of predicted targets do not exhibit altered expression. This could result from redundant gene regulation by other TFs or from false-positive GRN predictions. Collectively, these findings suggest that loss of function for single uncharacterized TFs might have limited phenotypic impacts but can reveal subsets of GRN predicted targets with altered expression.

Chapter III entitled ‘Transcriptome profiling of maize transcription factor mutants to probe gene regulatory network predictions’ is written and formatted as a manuscript. Some tables have been adapted and reformatted for thesis guidelines.

Erika L. Magnusson, Peng Zhou, Yi-Hsuan Chu, Peter Hermanson, Lina Gomez Cano, Zach Myers, Ankita Abnave, John Gray, Candice N. Hirsch, Erich Grotewold, Nathan M. Springer (est. 2023).

Several authors have made contributions to Chapter III work. Erika L. Magnusson, Peng Zhou, Yi-Hsuan Chu, Erich Grotewold, and Nathan M. Springer designed the experiments and interpreted the results. Erika L. Magnusson and Nathan Springer completed the majority of the writing. Candice N. Hirsch, Peng Zhou and Yi-Hsuan Chu helped with computational analyses. Lina Cano-Gomez, Ankita Abnave and John Gray provided data to include in this work. Peter Hermanson helped generate all the data for this manuscript.

CHAPTER III

Transcriptome profiling of maize transcription factor mutants to probe gene regulatory network predictions

Introduction

Transcription factors (TFs) transcriptionally regulate gene expression by recognizing and binding to DNA in a sequence-specific fashion. In eukaryotic genomes ~5-10% of the genes encode TFs that regulate transcription of all genes [136]. Gene regulatory networks (GRNs) represent the interactions between TFs and target genes that regulate spatial and temporal expression of a portion of all genes in an organism [70]. The data to match the discrete number of TFs to the larger number of target genes they regulate in GRNs remains limited. However, it is important to identify and understand how GRNs regulate endogenous metabolic pathways as this may provide a shortcut for modulating whole pathways or branch points of pathways [137]. Likewise, in crops, GRN inference could be used to select for existing variants or introduce new genetic variants [72] to rewire GRN architecture as a potential strategy to generate novel phenotypes for crop improvement.

Several methods have been used to predict TF-target gene interactions to generate GRNs in maize. These methods can include gene-centered approaches; a gene is known but regulators of the gene are not, or TF-centered approaches; the TF is known but the target genes it regulates are not [138]. Gene-centered approach methods can identify interactions where TFs directly bind to promoters or *cis*-regulatory elements (CREs) of a particular gene. Yeast-one-hybrid (Y1H) is a gene-centered approach that involves screening for interactions between the DNA sequence of interest (DNA bait) and a TF (protein prey) by activation of a reporter gene in yeast [138,139]. Studies in *Arabidopsis* [140–142] and maize [143] have successfully used Y1H to predict TF-target gene interactions. However, there are numerous limitations of the Y1H approach, including interactions are tested in yeast outside of genome tissue-specific or chromatin landscapes, cloned promoter sequences (DNA bait) are usually small (~1 kb) and may not capture the TF binding site, and interactions that require multiple TFs or post-translational modifications will be missed [139,144]. TF-centered approaches, such as chromatin-immunoprecipitation coupled with DNA sequencing (ChIP-seq), can identify potential TF targets genome-wide. ChIP-seq is a technique that uses a TF-specific antibody to selectively recover bound DNA from cross-linked DNA-protein complexes

[145,146]. In maize, direct targets for several TFs have been identified by ChIP-seq [147–150]. Unlike Y1H, ChIP-seq captures *in vivo* TF-target interactions within accessible chromatin regions and identifies target sequences that are both directly and indirectly bound (if the TF of interest forms a complex with another TF that directly binds). ChIP-seq is limited by the availability of antibodies and loss-of-function alleles to test antibody specificity. Another approach to GRN inference is to build gene co-expression networks (GCNs) with TF and target gene information and statistically measure the relationship between TF and target gene expression profiles. If the expression pattern of the TF and target gene are similar, these genes are considered co-expressed and may have shared regulation [151]. GCNs require large sets of quantitative data, usually RNA-seq, to capture gene expression and the statistical methods used for correlation can have a significant impact on the results [152]. Each approach to predict GRNs has some value, but comparisons among different approaches suggest many false positives and false negatives.

One method to test GRN predictions is to isolate loss-of-function mutants in TFs and test expression of predicted targets [78,79]. In maize, there are limited methods for moderate scale reverse genetics studies to assess if absence of the TF results in target genes with altered expression. Current maize mutant libraries only have functional knockouts for a subset of genes in the genome [153]. Further, the ability to test GRN predictions with TF loss-of-function alleles will vary based on the GRN prediction method. For example, the sample size of putative TF regulators identified by co-expression-based methods is usually much larger than those identified by Y1H screens. The number of Y1H predicted TF-target interactions is limited by the size and number of promoter regions that are cloned and tested, usually from a single putative pathway or functional type. GCN-based methods allow for construction of much larger networks and the likelihood of isolating loss-of-function alleles for some of these predicted TF regulators increases by the size of the network alone. Due to the recent whole genome duplication in maize [154], testing GRN predictions *in vivo* with functional TF knockouts may be limited by genetic redundancy with target gene regulation complemented by paralogs or members from the same TF family, requiring loss-of-function mutants for multiple related TF genes.

A set of putative loss-of-function alleles were recovered for 22 maize TFs with predicted targets based on either Y1H and/or co-expression-based networks. We did not observe phenotypic differences in these mutants. However, transcriptome profiling revealed variable numbers of differentially expressed (DE) genes in the mutants relative to wild-type plants. Gene ontology (GO) enrichment analyses identified functional groups of genes with altered expression in some of these mutants but did not point to clear biological functions for these TFs. The analysis of transcript abundance for GRN targets predicted using Y1H or co-expression analyses does reveal examples of predicted targets with altered expression. In some cases, the predicted targets are significantly over-represented in the DE genes (DEGs). However, the majority (> 75%) of predicted targets do not exhibit altered expression. These findings suggest limited perturbation of GRNs in single gene knockouts and could reflect high degrees of redundancy in gene regulation.

Results

To test the functional relevance of GRN predictions in maize we obtained stocks containing putative loss-of-function alleles for a series of TFs. Two primary sources of GRN predictions were utilized to select TFs for testing. The first source of GRN predictions was from a Y1H screen that identified putative TF regulators of maize phenolic biosynthesis [143]). This Y1H screen identified 45 TFs that exhibit interactions with at least 4 of the 54 phenolic biosynthesis gene promoters tested [143,155]. The second source of GRN predictions was a meta-analysis of TF-target gene co-expression from 45 GCNs [156]. To test the GRN predictions generated from GCNs, we identified 64 TFs that had ≥ 400 predicted targets and at least one coding sequence insertion indexed in the UniformMu mutant collection [22]. GRNs constructed from both sources, Y1H and GCNs, identified predictions for genes annotated in the B73v4 genome [157,158]. Before moving forward with network perturbation, all TFs selected for testing (45 from Y1H and 64 from co-expression) were confirmed to be single copy syntenic orthologs in the B73v4 and W22 genomes.

Isolation and characterization of TF loss-of-function alleles

Mutator (*Mu*) transposon insertions located within the set of TFs predicted from the Y1H or co-expression-based GRNs were identified using the sequence-indexed UniformMu

population created in a W22 inbred genetic background [22]. While most maize genes are associated with *Mu* insertions in this population, only 27.1% of annotated W22 genes have an insertion within the coding sequence [65]. We obtained all available insertions for the 45 TFs identified as candidates in the Y1H screen, including insertions in UTRs, introns, and proximal promoter regions. Given the prior evidence that some insertions within UTRs, introns, or 5' regions have minimal or no effect on the transcript produced by the allele [3,38,159], we focused primarily on coding region insertions, but for a few TF genes we also obtained 5' UTR, intron, and proximal promoter (up to 1 kb upstream of the TSS). We initially obtained stocks representing 150 alleles (82 TFs) but some of these were subsequently eliminated from our study because we could not confirm the presence of the insertion, because the insertion was not properly transmitted to subsequent generations, or lack of evidence for loss-of-function. This resulted in 32 putative loss-of-function alleles for 22 TFs including, 12 alleles for 8 of the TFs selected based on Y1H predictions and 20 alleles for 14 of the TFs selected based on GCN-based predictions (Figure 1, Table S1). None of the mutant lines used in this study exhibit notable morphological phenotypes (e.g., flowering time, plant architecture, or seed traits) that co-segregate with the mutation. It is possible that some lines have subtle phenotypic effects, but no major differences were observed over multiple field seasons.

Transcriptome profiling of TF mutant collection

Transcriptome profiling by RNA-seq was performed for each of the TF mutant alleles to characterize differential expression of the TF and the target genes. The TFs in the mutant collection exhibit variable tissue-specific patterns of expression based on data from 23 different tissues or developmental stages in B73v4 [117], so our experimental design focused on collecting mutant and wild-type transcriptome data from tissues with moderate-high expression of the TF relative to other tissues (Table S2). In total we surveyed five different tissues: coleoptile tip, imbibed embryo, seedling leaf, tassel, and tassel stem, with variable numbers of mutant alleles assessed in each tissue (Table S1, S2). To confirm the potential functional impact of the *Mu* insertion on the gene product of the TF, we assessed the expression level and transcript structure for each mutant allele. The $\log_2$ fold change in expression of each TF gene in the mutant allele harboring the *Mu* insertion relative to the wild-type allele was estimated from RNA-seq reads of all

biological replicates mapped to the W22 reference genome (which lacks the *Mu* insertion). This expression analysis revealed that only 8 of the 32 mutant alleles had significantly reduced expression levels and two (*wrky8-m1* and *hsf18-m1*) exhibited significant increases in total transcript abundance (Figure 2A, Table S3). However, the lack of a difference in expression level does not necessarily mean that a functional product is produced.

To identify potential variation in the transcript structure or sequence due to the *Mu* transposon insertion, we generated a *de novo* transcript assembly for each mutant and isolated the transcript structure produced by the mutant allele. We found that most mutant alleles have either altered transcript structure or sequence that is predicted to result in the production of truncated or altered protein sequences (Figure 2B, Table S4). In our study, we only retained mutant alleles that could not produce the full-length protein by selecting mutants with either transcripts 5' of the *Mu* insertion with the initial AUG out-of-frame and/or transcripts 3' of *Mu* that would only produce a protein less than half the length of the normal protein. Many of the assembled mutant transcripts reveal the presence of *Mu* transposon sequence. Mutant *de novo* assembled transcripts with *Mu* sequence near the annotated insertion site likely indicate that the *Mu* insertion is affecting mutant allele transcript structure. For several alleles there is evidence of two transcripts: one initiating at the gene promoter and terminating in the *Mu* sequence and the other initiating within the *Mu* element and reading through the remaining 3' sequence of the gene, and we could not detect amplified products using RT-PCR with primers flanking the *Mu* insertion site (Figure 2B, Table S4). Based on these assembled transcripts, it is unlikely that functional proteins are produced for most of these mutant alleles, although it is possible that partial fragments could be generated in some cases.

The overall changes to the transcriptome were assessed through principal component analysis for each tissue (Figure S2). In general, the samples tended to cluster by mutant allele, but some samples were closely similar to W22 wild-type while others were more distinct (Figure S2). Genes that were significantly DE were identified for each mutant allele relative to wild-type W22 replicates from the same tissue (Figure 3). The number of DEGs was quite variable with some mutants only exhibiting ~100 DEGs and others having > 1,500 DEGs (Figure 3). For these putative TFs it is not known whether they

function as primarily activators or repressors; therefore, both up- and down-regulated genes were identified for each mutant allele. Most of the mutants (26/32) have more down-regulated genes than up-regulated genes, which would be predictive of primary targets of transcriptional activators. We examined the 10 TFs that are represented by two independent mutant alleles and found significant (hypergeometric enrichment $p < 0.05$) overlap in both the up- and down-regulated DEGs for all 10 pairs of mutant alleles. While the overlap of DEGs between the two mutant alleles was highly significant, the proportion of the up- or down-regulated genes that are significant in both mutant alleles was highly variable (Figure S3). In several cases > 50% of the DEGs are identified in both mutant alleles, but in other cases the overlap only accounted for 5-10% of the DEGs (Figure S3).

TF mutant allele DE genes reveal enriched GO terms

The potential functional role of the TFs as regulators was investigated by monitoring enrichment for gene-ontology (GO) terms in DE genes [160]. There are a relatively large number of terms with statistical significance for each mutant allele. To compare potential functional enrichments for the different mutants, we identified the non-redundant set of 50 GO terms with the most significant enrichment for up- or down-regulated DEGs among all 32 mutant alleles (Figure 4, S4). There are several examples in which the two independent mutant alleles for a TF or mutant alleles for TFs in the same family exhibit consistent enrichments. For example, the two mutant alleles of *E2F19* exhibit down-regulation enrichment of methyl esterase activity and UDP-glycosyltransferase (UGT) activity (Figure 4). The two mutant alleles of *WRKY87* and one allele for *WRKY82* all exhibit enrichment for down-regulation of mini-chromosome maintenance (MCM) and THO complex genes (Figure 4). There were several GO terms that exhibit significant enrichment in at least eight of the mutants including, UDP-glycosyltransferase (UGT), THO complex, MCM complex, DNA unwinding, and cellular glucan metabolism. While some mutants have functional enrichments for DEGs, the uncertainty in identifying the functional role of the TFs based on the annotation of DE genes is consistent with plants that display no major phenotypic differences under standard growth conditions.

A small number of Y1H predicted targets exhibit differential expression

Only a subset of the TFs (13/22) with mutant alleles have targets predicted from the Y1H analysis. These include five TFs with three or less predicted targets and eight TFs with at least four predicted targets (Figure S5). There are some predicted Y1H targets that are not expressed or lacking a W22 gene annotation (Y1H targets were predicted for B73v4 based gene annotations) in either the mutant or wild-type plants (Figure 5, S6) and these could not be tested for altered expression in the mutants. The majority of the predicted targets that were expressed did not show evidence for differential expression in the mutants (Figure 5, S6). There were only four significantly down-regulated and seven significantly up-regulated Y1H predicted target genes out of more than 100 predicted targets tested for differential expression. Given the relatively small number of targets for each mutant allele, it is difficult to perform a formal significance analysis, but our results provide some initial evidence that removing a single TF has minimal functional consequences for the expression of most Y1H predicted targets. The analysis of multiple alleles for the same TF did reveal one potentially interesting case of confirmed functional effects for Y1H predicted targets. The gene *Bz1* (*Zm00001d045055;* a UGT involved in the glycosylation of anthocyanidins) is a predicted target of *HSF24* and is significantly down-regulated in both *hsf24-m3* and *hsf24-m4* (Figure 5, S7). This gene contains two predicted HSF binding sites located in the Y1H cloned promoter region. The *E2F19* predicted target gene *HCT11* (*Zm00001d020530*) is significantly down-regulated in *e2f19-m1* and has decreased expression but is not significantly DE in *e2f19-m2* (Figure S7). The *MYB40* predicted target gene *A1* (*Zm00001d044122*) has five MYB transcription factor binding sites within the Y1H cloned promoter region and is significantly up-regulated in *myb40-m1*. This gene is not significantly changed in *myb40-m2,* but it does show increased expression in two of the three biological replicates for this allele (Figure S7).

Some TF mutants exhibit enriched differential expression of co-expression-based GRN predicted targets

The co-expression-based GRN predictions provide a much larger set of predicted targets for most of the TFs in this study (Figure S5). Our analysis primarily focused on the full set of predicted targets that are identified in at least one ($n \geq 1$, referred to as n1) of the GCNs, but for some analyses we also assessed enrichments for targets that were

found in at least three (n ≥ 3, referred to as n3) of the GCNs. We observed 13 out of the 29 alleles with GRN predicted targets exhibit significant enrichment (Figure 6A). There are three mutant alleles with significant enrichment for only down-regulation, four mutant alleles with significant enrichment for only up-regulation, and 6 mutant alleles (*hsf18-m1*, *hsf20-m1*, *hsf6-m1*, *mybr32-m*1 and *wrky82-m1*) with significant enrichment for both up- and down-regulated target genes (Figure 6A). In most cases, the enrichment of targets is only 1.5 to 2-fold, (Figure 6A). While there is a significant enrichment for differential expression of predicted target genes for some of the TFs, the majority (> 80%) of the GCN-based GRN predicted targets do not exhibit differential expression in the mutant relative to wild-type (Figure 6B). A similar set of analyses were performed after restricting the predicted targets to genes identified in at least three of the GCN-based GRNs, GRN n3 (Figure S8). For several of the TFs, the targets predicted in at least three GCN-based GRNs, GRN n3, exhibit higher enrichment for differential expression than those predicted in GRN n1 (Figure S8A, Figure 6A), but most predicted targets still do not exhibit altered expression (Figure S8B).

Discussion

Perturbing GRNs is considered a potential mechanism to influence traits in many species. Substantial investments have been made in developing GRNs to predict the functional targets for many TFs and to generate mutant collections for maize genes. In this study, we have monitored the consequences of mutant alleles for 22 maize TFs that have predicted targets. These TFs were selected based on prior Y1H work to identify TFs that bind to promoters of multiple genes in the phenylpropanoid pathway [143] or based on co-expression network analyses [161]. All the mutant lines that were analyzed are derived from the UniformMu population and are in a W22 genetic background [67]. Each mutant line was grown in at least two field seasons with replicated plots. No obvious morphological differences were observed in segregating rows comparing mutant and wild-type siblings or in plots of homozygous mutants compared to standard W22. There were no obvious differences in flowering time or ear traits. It is worth noting that in reverse genetics studies it is much easier to evaluate specific predicted traits rather than to search for any potential phenotype. However, it was clear that the loss-of-function for the single TFs used in this study did not result in major phenotypes in the field environments.

Transcriptome profiling was utilized for three distinct purposes in this study: evaluation of the mutant allele transcripts, GO enrichment analyses of genes with altered expression, and assessment of expression changes at GRN predicted targets. We initially performed transcriptome profiling on a larger set of UniformMu mutant alleles; however, for this study we focused on the subset of alleles that most likely represent loss-of-function mutations. Evaluation of TF gene expression levels in the mutant alleles revealed some cases of reduced expression, but most genes did not exhibit a significant change in transcript abundance. A more detailed analysis of coverage based on the RNA-seq data found that most mutant alleles had similar patterns of read depth coverage compared to W22 in regions of the gene both 5' and 3' of the *Mu* insertion. However, the regions near the *Mu* insertion often had reduced coverage, as might be expected given the sequence differences in this region, which would reduce the ability to map short reads to the reference genome. To better understand the transcript structure in the mutant alleles, we performed *de novo* transcriptome assembly. We found that in these mutant alleles, which mostly represent coding sequence insertions, there are often multiple transcripts. One of the transcripts represents the 5' portion of the gene with the expected 5' UTR and splicing, up to the site of the *Mu* insertion. This 5' transcript would include some sequence of the *Mu* terminal inverted repeat and then truncate within the *Mu* element. *In silico* translation of such transcripts suggested the potential to make a fragment of the wild-type protein. The other transcript appeared to initiate within the *Mu* element and then read through to the expected transcription termination site. This is reminiscent of *Mu*-suppressible alleles that have been previously characterized. *Mu*-suppressible alleles occur when an outward-reading promoter in the *Mu* element initiates transcription in plants with epigenetic silencing of *Mu* transposition [49,53]. Many of the previously characterized suppressible *Mu* alleles represent insertions in proximal promoters or 5' UTRs that can produce the full-length protein. The highly significant overlap of DE genes (hypergeometric enrichment $p < 0.05$) between the two independent alleles of six of the TFs support the conclusion that the materials used in this study represent actual loss-of-function alleles.

The evaluation of enrichment in GO terms can be useful in understanding functional consequences of the mutant alleles. Many of the mutants exhibited significant enrichments for some GO terms in the up- or down-regulated genes. However, we found

it difficult to pinpoint a specific function of the TFs based on these GO terms. A similar analysis was also performed using the maize metabolic pathways [162], and while there were some pathways with significant enrichments, there was no clear evidence that any of our TF mutants exhibited broad, consistent changes to transcripts associated with specific pathways.

The evaluation of transcript abundance for genes that are predicted as targets for each of the TFs revealed some examples of altered expression. In about a third of the TFs with co-expression-based GRN predictions there was a significant enrichment of the predicted targets within the DEGs. However, the majority of the DEGs are not predicted targets and the majority of the predicted targets are not DE. We considered two main explanations for this observation. One potential explanation is that the GRN predictions have a high rate of false positives. The types of data used to generate the GRN predictions used in this study are both known to have false positives. Y1H assays are conducted in the absence of the normal chromatin environment of the endogenous promoters and can generate false positives. Co-expression analyses are simple guilt-by-association approaches and can suggest functional interactions for sets of genes that are co-expressed in similar patterns, even if these genes have independent regulation. While these false positives can occur in both types of GRN predictions, we are not confident that this is a primary explanation for the low validation rates we observed. A second explanation is the potential for genetic redundancy in the regulation of these predicted targets. This redundancy could be due to highly similar TFs (either retained duplicated genes from the recent whole-genome duplication event in maize or other members of the same TF family), or due to other TFs that independently regulate the same target gene. It is worth noting that, in many cases, there are substantial numbers of DE genes in the single TF knockouts, so there must be some non-redundant function of these genes along with the possibility that there is partially redundant regulation. The explanation of redundancy could explain why some of the predicted targets are not DE, but in many cases, we did not find evidence from Y1H or co-expression data that would have implicated other highly similar TFs. This might suggest that the GRN predictions tend to highlight specific potential TF-target interactions when the biological reality might be much more complex. In most cases we were not able to recover loss-of-function alleles for multiple related TFs which reduces our ability to perform targeted analyses of

redundant regulation. Future studies that utilize genome editing or other approaches to create loss-of-function alleles in multiple TF family members could better explore the potential redundancy of regulation in these pathways.

Methods

Isolation of TF mutant alleles from UniformMu populations

UniformMu transposon-indexed seed stocks were ordered from the MaizeGDB Stock Center [116] (Table S1). Seeds were planted at field sites on the Saint Paul campus to isolate DNA for genotyping from Vegetative 3 (V3) leaf tissue and to maintain stocks by selfing. Mutant alleles were genotyped to identify the presence and zygosity of *Mu* with gene-specific primers flanking the *Mu* insertion and primer with specificity to *Mu* TIR regions: 9242 (Table S1) [22]. Homozygous transmissible alleles were then isolated after backcrossing twice to the W22 r-g (colorless) inbred if possible, to reduce the original mutation load from the transposon-indexed stock (Table S1). Assessment of phenotype were performed for multiple plots and multiple growing seasons for each of the mutants.

Plant material and RNA-seq experimental design

Homozygous transmissible alleles with various pedigrees (Table S1) were grouped for RNA-seq sampling by tissue where the TF gene has moderate-high expression from previously published expression profiling data in B73 [117] and W22 [118] (Table S2). In total, five different tissues were selected to sample to capture expression of each of the 22 TF genes in mutant and control conditions: coleoptile tip, seedling leaf, imbibed embryo, tassel, and tassel stem. Three biological replicates of the mutant alleles and at least three biological replicates of control W22 *r-g* (colorless) in each tissue were sampled for RNA-seq (Table S1, S3). Samples for each tissue were collected on the same day at the same time with the 24 h time sampled listed in parentheses. At anthesis, tassel stems (~3 cm) and whole tassels (anthers unextruded) were sampled (9:00) from 3 plants in the field each and pooled for one biological replicate. Embryos were dissected (from 8:00-11:00) after imbibing seeds in distilled water for 48 h at 31°C and 5 embryos were pooled for one biological replicate. For seedling leaf tissue, the newest collared leaf was sampled (9:00) from each seedling 10 DAS in 16 h light 28°C, 8 h dark 24°C growth chamber conditions and 3 leaves were pooled for one biological replicate. Coleoptile tips were sampled (9:00) from seeds 6 DAS in 30°C dark conditions using a paper towel

cigar roll method for germination [119] and 3 tips (~2.5 cm) were pooled for each biological replicate.

RNA-seq data processing

Total RNA was extracted using Qiagen RNeasy Plant Mini Kit (QIAGEN, Inc., Valencia, CA) and quantified internally and externally by the University of Minnesota Genomics Center (UMGC) with the Quant-iT RiboGreen RNA Assay Kit (Thermo Fisher, Cat. No. R11490) and quality checked by UMGC with the Agilent 2100 Bioanalyzer. For one mutant allele, *gras52-m1*, one biological replicate sample did not meet the minimum RNA concentration for RNA-seq library preparation and was discarded prior to sequencing. Sequence libraries were prepared using the standard TruSeq Stranded mRNA library protocol and sequenced on the NovaSeq 150 bp paired end S4 flow cell to produce at least 20 million reads for each sample. For all samples with paired end sequencing, both library construction and sequencing were done at UMGC. Library construction and sequencing for two mutant alleles, *hsf24-m3* and *hsf24-m4*, and W22 control sampled from tassel tissue was done externally at the Genomic Core at Michigan State University. For these samples, libraries were prepared from 2 µg of total RNA using the TruSeq RNA Sample Prep Kit (Illumina, Cat # FC-122-1001) and sequenced on the HiSeq 4000 to produce at least 18 million 50 bp single-end reads.

For all samples, sequencing reads were then processed through the nf-core RNA-Seq pipeline [120,121] built with Nextflow v20.10.0 [121] for initial QC and raw read counting. Reads were trimmed using Trim Galore! v0.6.5 [122] and aligned to the W22 reference genome [65] using Hisat2 v2.1.0 [123] with default parameters (“hisat2 -x $db $input -p 12 --met-stderr --new-summary”). Uniquely aligned reads were then counted per feature by featureCounts v2.0.1 [124]. Raw read counts were then normalized by library size and corrected for library composition bias using the TMM normalization approach in edgeR v3.28.0 [125], to give CPMs (Counts Per Million reads) for each gene in each sample, allowing direct comparison between mutant and control samples. CPM values were then normalized by gene CDS lengths to give FPKM (Fragments Per Kilobase of exon per Million reads) values. Genes were considered expressed if their CPM ≥ 1 in at least one sample per tissue. Principal component analysis of expressed genes CPM values were used to

explore sample cluster patterns among and between biological replicates of mutant alleles and wild-type W22 control.

Transcriptome profiling

Reads from RNA-seq data of combined biological replicates for each allele, mutant or control, were *de novo* assembled into transcripts with TRINITY v2.5.1 [127]; Reads were quality trimmed with Trimmomatic v0.33 [128] and strand-specificity was not defined. A local blast database (SequenceServer [129]) was created for each *de novo* transcriptome assembly to identify transcripts aligning to the W22 TF gene cDNA in each mutant and control. W22 control transcript assemblies for each gene were first analyzed by both BLASTn [130] and the ExPASy translate tool [131] to confirm TRINITY could assemble the full-length gene cDNA from the RNA-seq short-read data. The canonical ORF of each TF gene was identified by comparing the annotated W22 TF gene cDNA sequence to sequences of orthologous genes in other grass species (i.e., *Sorghum Bicolor, Setaria Italica, Oryza sativa*) via BLASTx. Mutant allele transcripts were further analyzed by BLASTn to determine the effect of the transposon insertion and identify if *Mu* sequence was transcribed.

Identification of TF mutant allele DE genes

Raw read counts of genes from all replicates of each TF mutant allele and W22 control from the same tissue were used to call DE genes: false discovery rate [FDR] adjusted p-value < 0.05 and a minimum fold change of 2 (DESeq2 v1.30.1 [126]). To determine if any genes could be DE due to genetic differences between the control W22 (r-g) colorless and UniformMu color-converted W22 [22], DE genes were called between all TF mutant allele and W22 samples from the same tissue. A small number of genes (n = 1773) that are consistently DE in more than half of the mutant alleles from imbibed embryo, tassel stem and seedling leaf tissues (< 50% of mutant reads are from one TF gene) were removed. DE genes for each TF mutant allele were then filtered for expressed genes (CPM ≥ 1 in at least one sample per tissue) and genes with one-to-one gene models between W22 [65] and B73v4 genomes [157,158]. Filtering for W22 mapped genes with one-to-one B73v4 gene models was necessary for testing enrichment of predicted B73 mapped co-expression-based GRN and Y1H TF targets.

Identifying TF binding sites in Y1H promoter cloned regions

Transcription factor binding sites (TFBSs) were identified in Y1H bait cloned promoter sequences of predicted targets [143] for two TFs, *MYB40* and *HSF24*. The B73v4 maize genome was scanned for TFBS sequence of five R2R3 MYB TFs (*AtMYB52*, *AtMYB59*, *AtMYB46*, *AtMYB111*, *AtMYB55*) [163] and a generic HSF TFBS (5' *NGAANNTTCN* 3') [164], with N nucleotide weighting set to the GC content of maize genomic DNA to find significant matches ($p < 0.01$) (FIMO tool in the MEME suite [165]). These putative genome-wide TFBSs were then subset to Y1H bait promoter cloned regions for predicted targets of MYB40 or HSF24 to identify if these promoter sequences contained the respective TFBSs (BEDTools intersect [133]).

Enrichment for shared DEGs between multiple independent alleles per TF

A significant hypergeometric enrichment of finding more than the expected number of overlapping DEGs between two independent mutant alleles per TF required a representation factor > 1 and $p < 0.05$ [166]: expected number of genes = (number of genes in allele A × number of genes in allele B) / number of expressed genes. The number of possible shared DEGs between two independent mutant alleles per TF is the minimum number of either up- or down-regulated DEGs between both alleles. The number of shared DEGs was calculated as a proportion out of the possible shared DEGs between two independent alleles per TF in Figure S3.

Enrichment for TF predicted targets

All statistical analyses for enrichment of DEGs utilized a hypergeometric probability. Testing for over-representation or enrichment of mutant allele DEGs for GO terms or GRN predicted targets was calculated with R stats v4.0.2 hypergeometric phyper(q = x - 1, m, n = N - m, k, lower.tail = FALSE) function, $p < 0.05$ for significance ([134]. For GO term enrichment: N = number of expressed genes associated with any GO term, m = number of genes with a specific GO term, k = number of DEGs with any GO term, x = number of DEGs with a specific GO term. For GRN (n1 and n3) predicted target enrichment: N = number of expressed non-redundant predicted target genes in the genome, m = number of expressed predicted targets per TF, k = number of DEGs, x = number of DEGs that are predicted targets. The fold-enrichment of GRN predicted targets that were DE was calculated by [x / (k / N) × m].

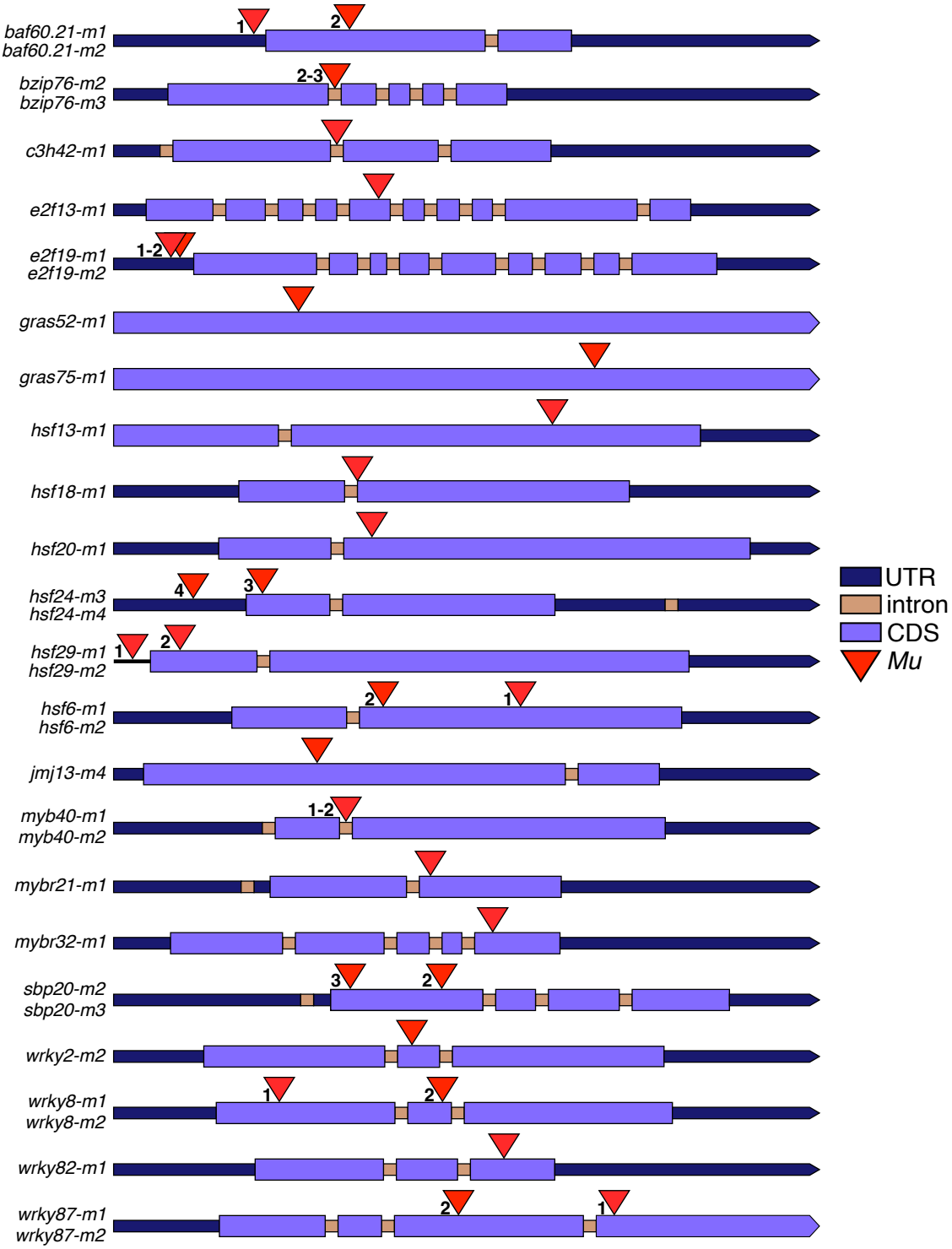


Figure 1. Schematic of 32 maize UniformMu mutant alleles isolated in 22 transcription factor genes. TF gene model annotation features are indicated by different colors/shapes in the plot. Independent mutant alleles from the same gene are depicted by numbers that correspond to the unique identifier of each allele. *Mu* transposons vary in insertion position within the annotated gene. Each gene model UTRs and CDS' are scaled proportionally, but introns are not shown to scale. All genes are shown in the forward orientation for plotting purposes.

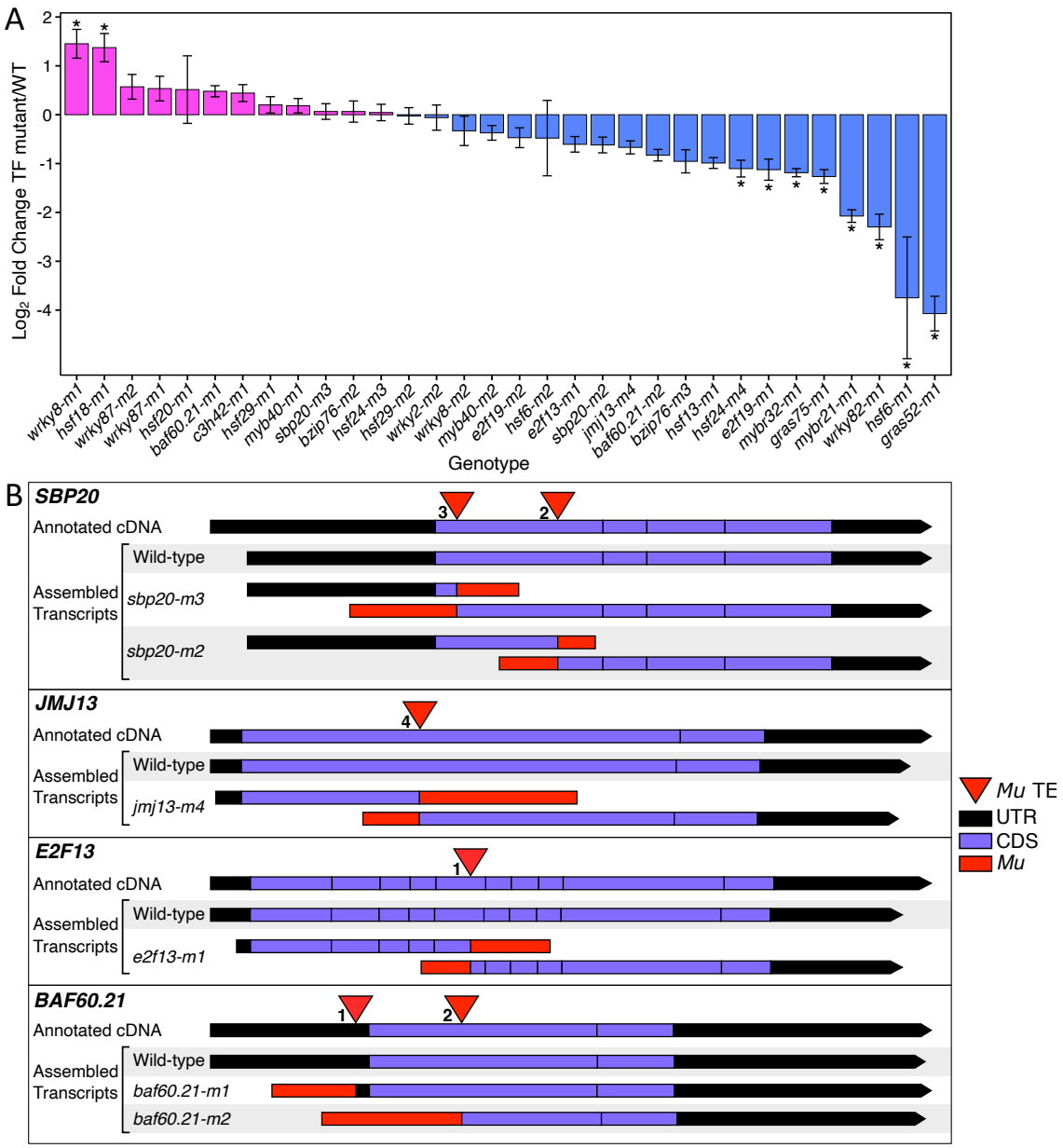


Figure 2. Changes in transcript abundance and structure for TF mutant alleles. (A) The fold change in expression of each TF gene in the mutant allele harboring the *Mu* insertion relative to wild-type plants was estimated from RNA-seq reads of all biological replicates mapped to the W22 reference genome (Log_2 Fold Change TF mutant allele/WT). The standard error of the Log_2 fold change (Log_2fc) estimate is represented as error bars for each allele. Significant differential expression between the TF mutant and W22 control is indicated by an asterisk (*). Alleles are rank ordered based on the difference in expression for this plot with positive Log_2fc: pink and negative Log_2fc: blue. **(B)** Schematic of *de novo* assembled transcript structure(s) for mutant allele and W22 WT cDNA relative to annotated WT cDNA. Assembled transcripts for mutant alleles: *sbp20-m2*, *sbp20-m3*, *jmj13-m4* and *e2f13-m1* show a split gene model with one transcript initiating within *Mu* and a second transcript initiating from the native TSS with a premature termination in *Mu*. *baf60.21-m1* and *baf60.21-m2* have one transcript assembled initiating within *Mu.* All *de novo* transcripts were assembled with TRINITY.

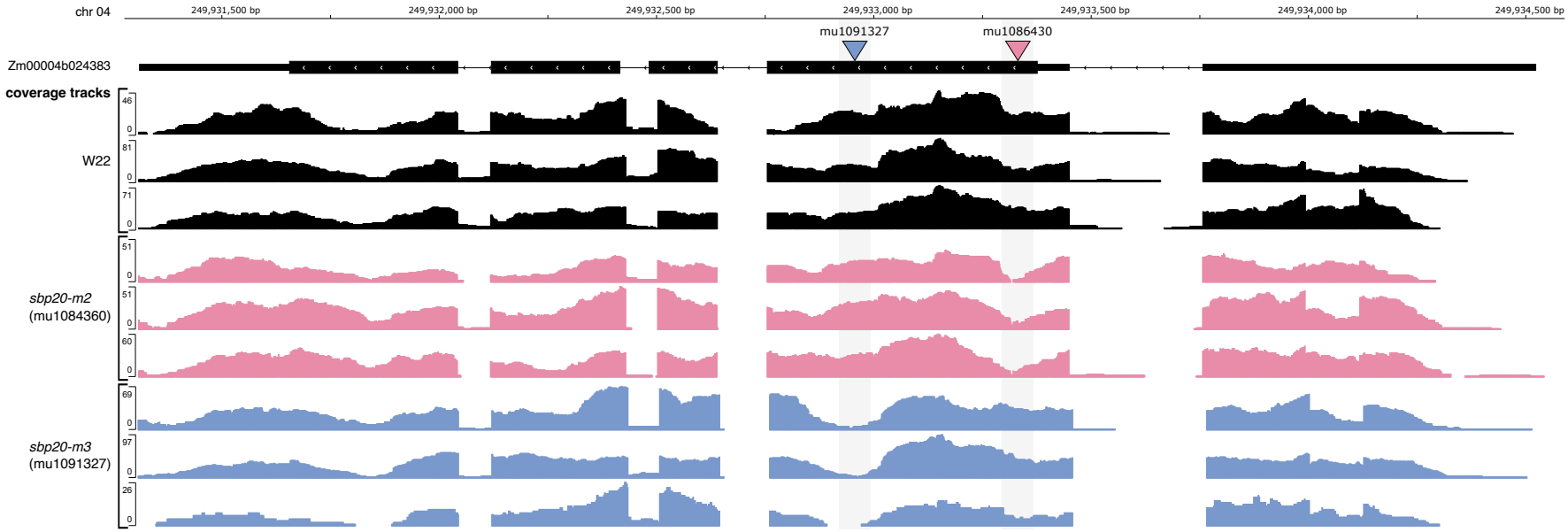


Figure S1. Visualization of RNA-seq read coverage for *Sbp20 (Zm00004b024383)* wild-type and mutant alleles; *sbp20-m2* (mu1084360) and *sbp20-m3* (mu1091327). IGV was utilized to visualize the coverage of RNA-seq reads for each biological replicate. The *sbp20-m2* (pink) *and sbp20-m3* (blue) mutant allele coverage tracks both show reduced coverage in the regions flanking their respective annotated *Mu* insertions (triangles) in the mutant samples compared to W22 wild-type.

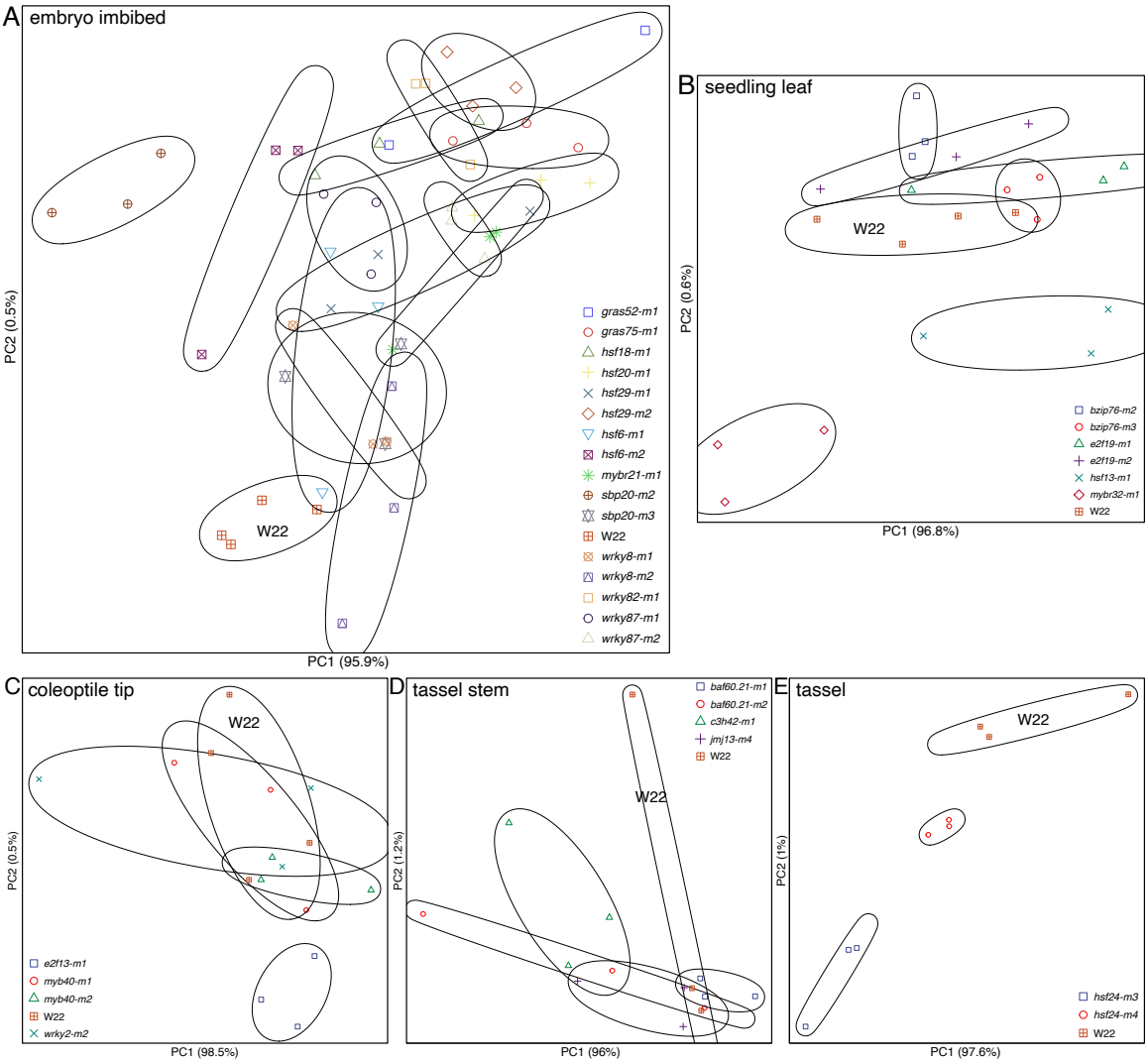


Figure S2. Principal component analysis (PCA) of RNA-seq data. PCA was performed separately on the RNA-seq samples for each of the 5 tissues: **(A)** embryo imbibed, **(B)** seedling leaf, **(C)** coleoptile tip, **(D)** tassel stem, **(E)** tassel. Different symbols/colors are used for each genotype and ellipses are used to show the group of three biological replicates (except for *gras52-m1* which only has 2 replicates). Clustering was performed using CPM values for all genes that are expressed at > 1 CPM in control samples in the relevant tissue. Gene CPM values were arcsine transformed and scaled to have unit variance. The base R function prcomp() was used to calculate principal components.

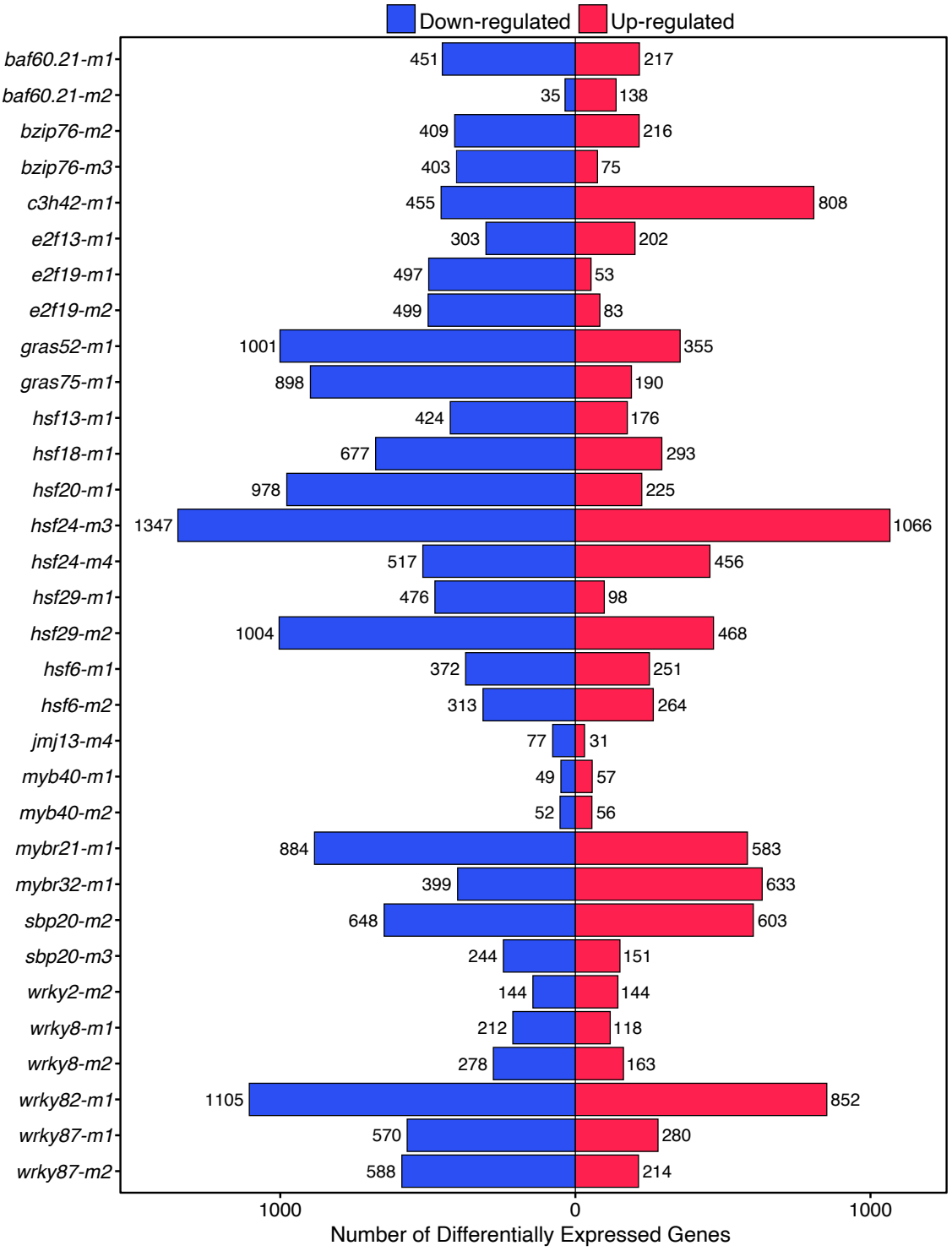


Figure 3. Identification of differentially expressed genes in transcription factor mutant genotypes. Genes with significant differences in expression were determined for each mutant based on comparison to W22 samples of the same tissue type using DEseq2 (p-adjusted <0.05 and >2 fold-change). The number of significantly up- (red) or down-regulated (blue) differentially expressed genes (DEGs) in the TF mutant allele relative to the W22 control is shown.

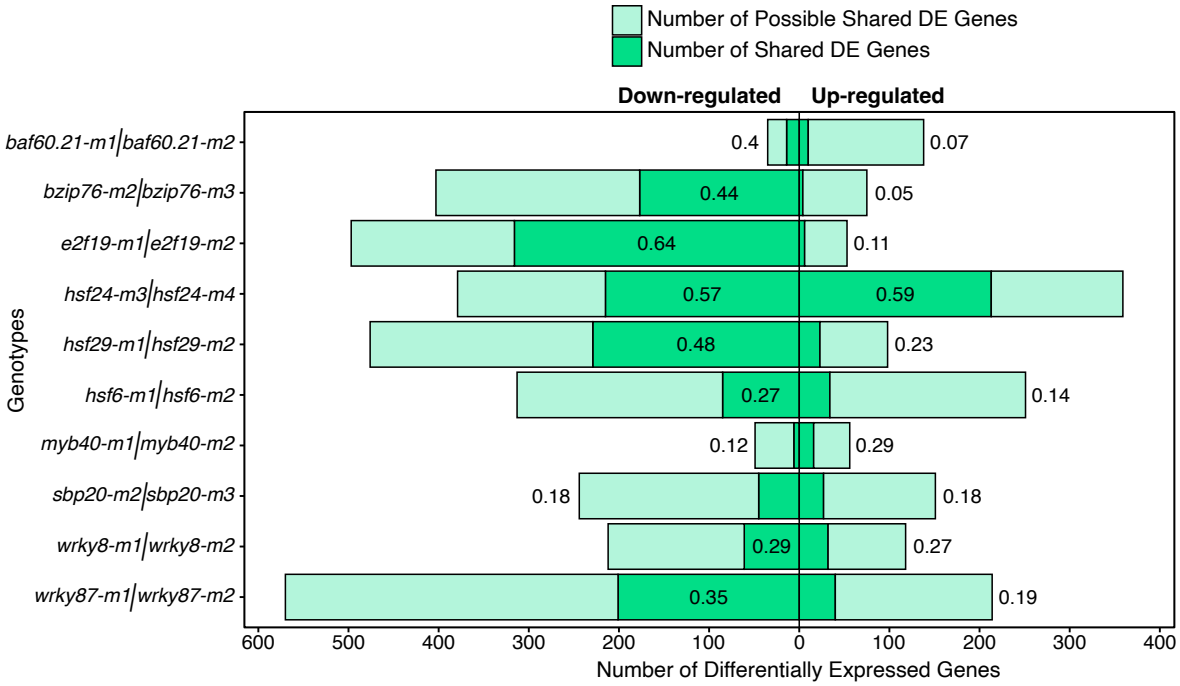


Figure S3. Proportion of shared out of possible down- or up-regulated differentially expressed (DE) genes for transcription factors with multiple independent mutant alleles. The number shared DE genes (dark green) as a proportion of the total number of possible DE genes that could be shared between alleles; minimum number of DE genes between alleles (light green). Proportion calculated is shown within the shared (dark green) bar or adjacent to the bar for each TF and direction. For all 10 TFs, the proportion of both down- and up-regulated DEGs shared between the two mutant independent alleles per TF (dark green) represents a significant hypergeometric enrichment ($p < 0.05$) of finding more than the expected number of shared DEGs.

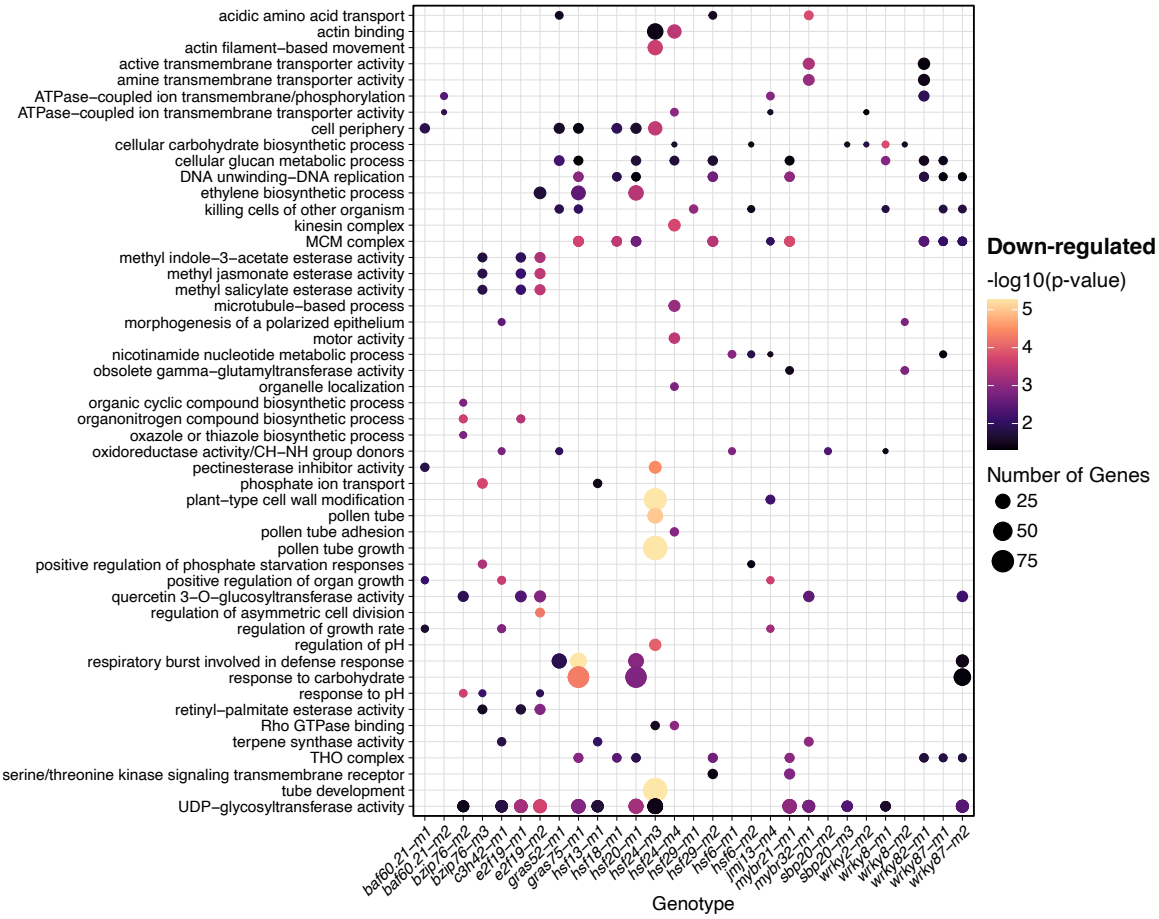


Figure 4. GO-based analysis of TF mutant down-regulated differential expression. A GO analysis of the down-regulated genes for all 32 mutants was used to identify a set of the top 50 non-redundant enriched terms (lowest p-values). The GO enrichment levels were determined by a hypergeometric test, where the GO term observed/expected number for each gene set was tested. Circles are used to indicate each significant enrichment of GO terms for 29 of the 32 mutants (*e2f13-m1, myb40-m1*, and *myb40-m2* have no significant enrichments in the top 50 GO terms for down-regulated DEGs).

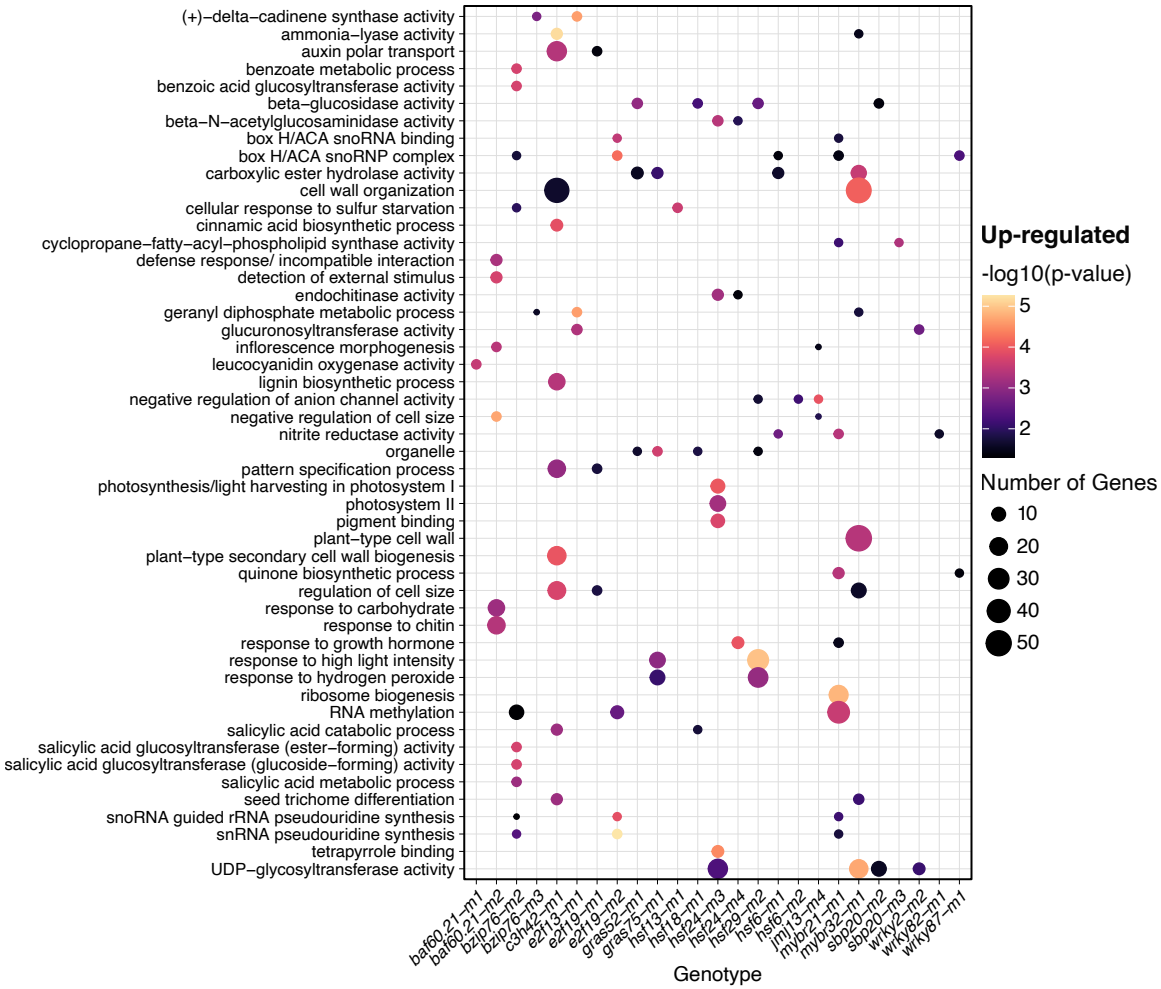


Figure S4. GO-based analysis of TF mutant up-regulated differential expression. A GO analysis of the up-regulated genes for all 32 mutants was used to identify a set of the top 50 non-redundant enriched terms (lowest p-values). The GO enrichment levels were determined by a hypergeometric test, where the GO term observed/expected number for each gene set was tested. Circles are used to indicate each significant enrichment of GO terms for 25 of the 32 mutants (*hsf20-m1, hsf29-m1, myb40-m1, myb40-m2, wky8-m1, wrky8-m2, wrky87-m2* have no significant enrichments in the top 50 GO terms for up-regulated DEGs).

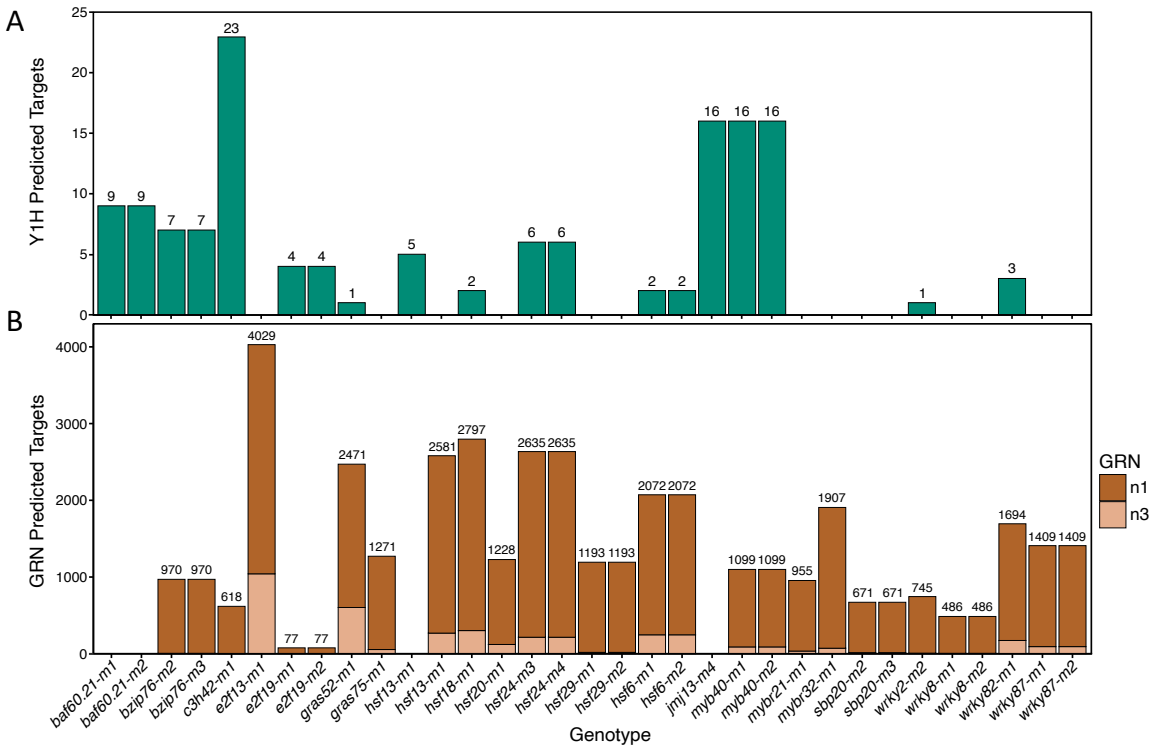


Figure S5. The number of predicted targets for each TF gene with mutant alleles. Prior GRN work using yeast one-hybrid (Y1H) or co-expression approaches have predicted putative GRN targets for many maize TFs. For the 22 TFs assessed in this study we show the number of predicted targets based on Y1H **(A)** or co-expression based GRNs **(B)** with mutant alleles tested from GRN co-expression and pathway-based methods. These values represent the number of predicted targets based on analysis of B73v4 gene annotations. The Y1H targets include genes identified in Yang et al. (2017) as well as additional targets based on further screening. The co-expression based GRN targets are derived from Zhou et al., (2019) and were filtered to require syntenic orthologs annotations in W22 and expression of CPM > 1 in the tested tissue. Tan (n3) indicates co-expression based GRN targets identified in at least 3/45 co-expression GRNs while brown (n1) indicates targets detected in at least one of the 45 GRNs.

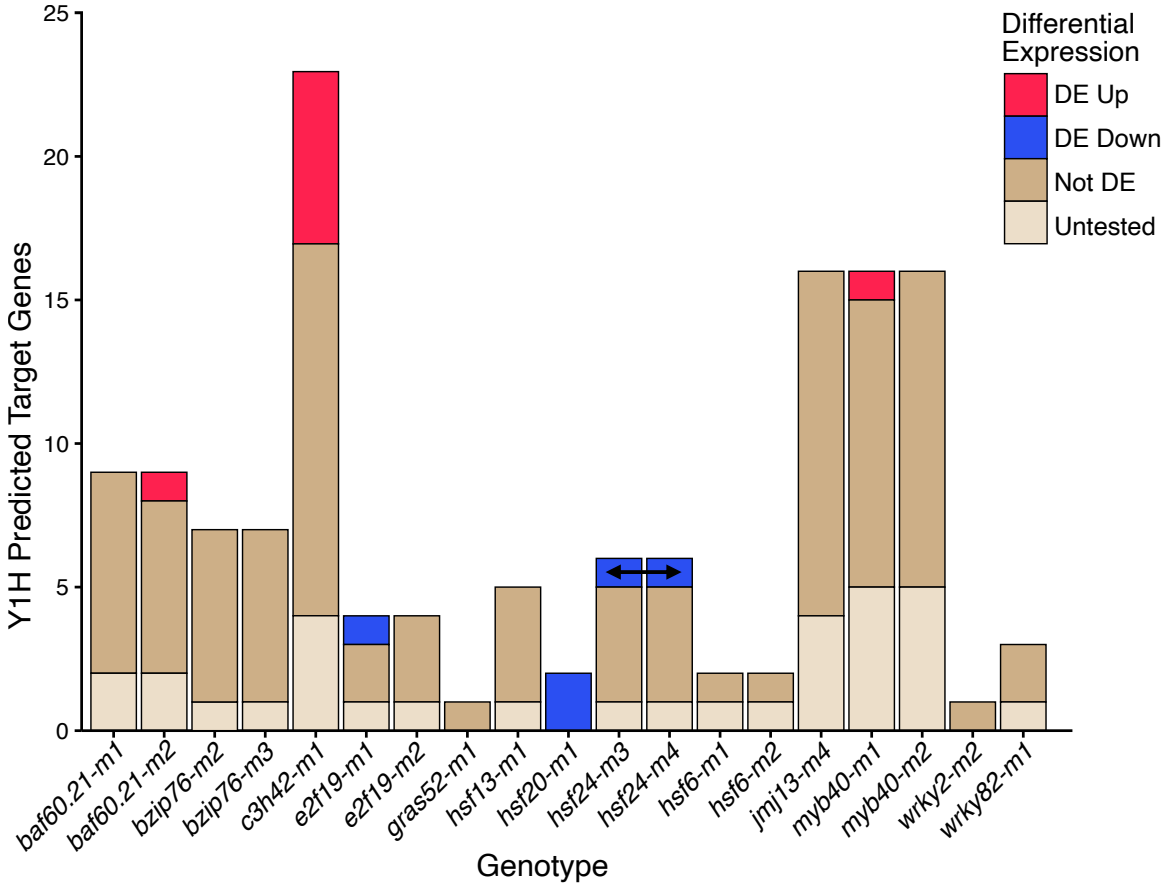


Figure 5. Limited examples of differential gene expression for targets predicted by yeast one-hybrid. The expression of each of the putative targets of each TF that were identified in Y1H analyses (Table S5) was assessed in the mutants. Each predicted target was classified as DE (Up or Down), not DE or untested. The untested genes include genes that did not have an annotation in the W22 genome and genes that are not expressed in this tissue. The predicted target (*Bz1*, Zm00004b031704) that is significantly down-regulated in both *hsf24* mutant alleles is indicated by an arrow.

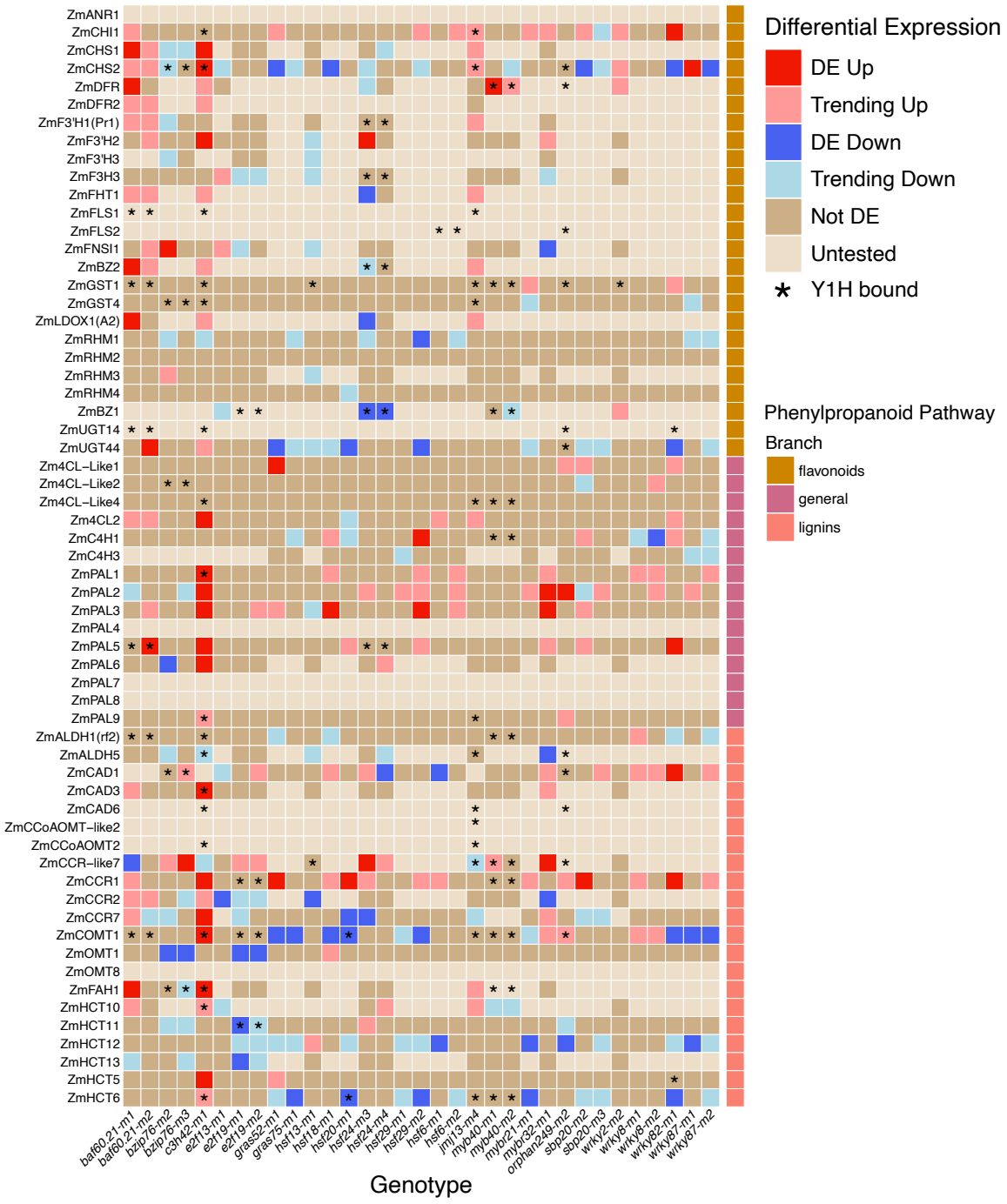


Figure S6. Expression changes for phenylpropanoid pathway genes in the TF mutants. The relative expression in mutant compared to wild-type is shown for 61 phenylpropanoid pathway gene Y1H targets. The phenylpropanoid pathway genes are ordered by pathway branch (general, flavonoids, lignins) and enzyme with color coding on the right side of the plot. Y1H positive interactions are shown using an asterisk (*). The relative expression is indicated using color [DE Up or Down: differentially expressed (log2 fold change ≥ 1 and FDR adjusted p-value < 0.05); Trending Up or Down: fold change ≥ 1.5, but not differentially expressed; Not DE: not differentially expressed and fold change < 1.5; Untested: not expressed (CPM < 1) or no W22 gene annotated].

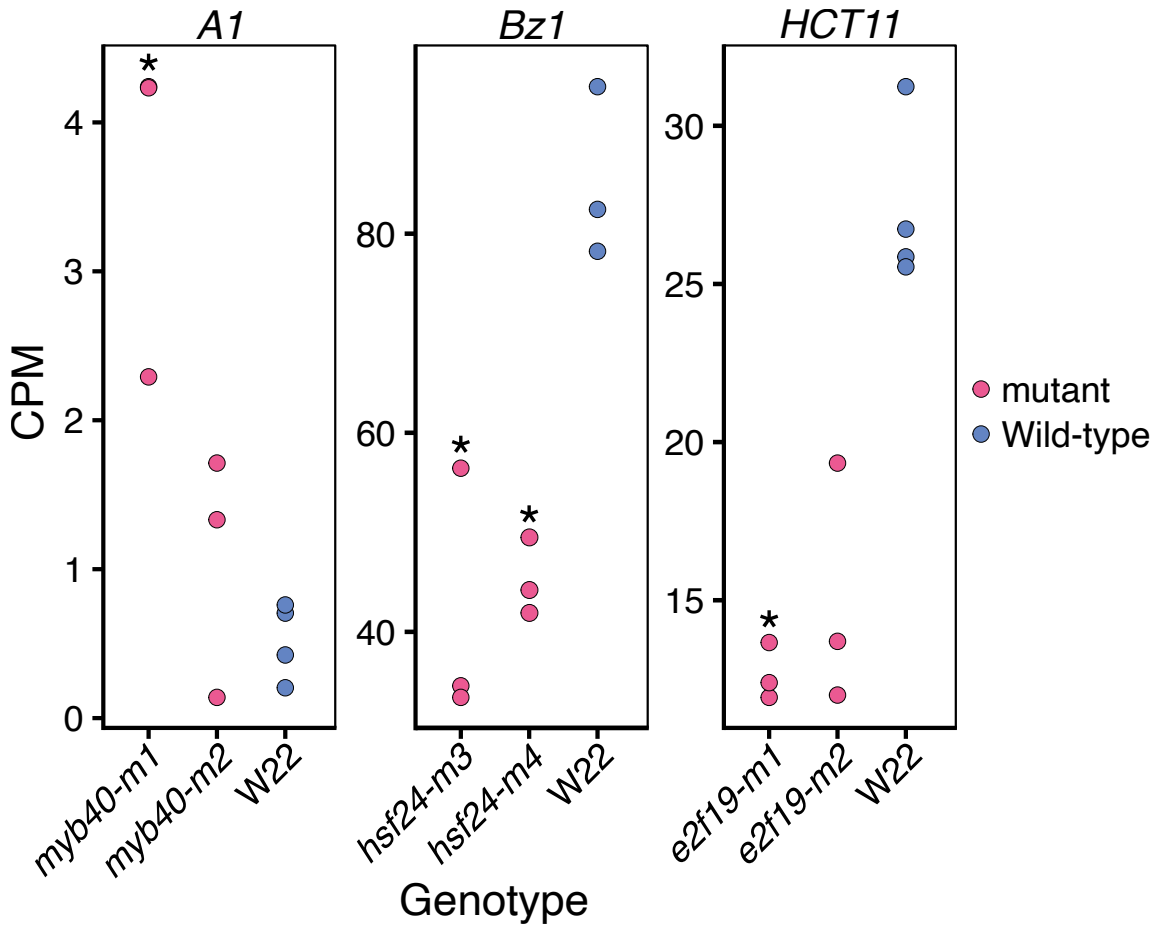


Figure S7. Expression differences for several Y1H predicted targets. The CPM values for each of the biological replicates are for shown for three Y1H phenolic target genes: *MYB40-A1, HSF24-Bz1, and E2F19-HCT11*. Mutant allele (pink) and wild-type W22 (blue) biological replicates are plotted comparatively with significant differential expression indicated (*); log2 fold change ≥ 1 and FDR adjusted p-value < 0.05.

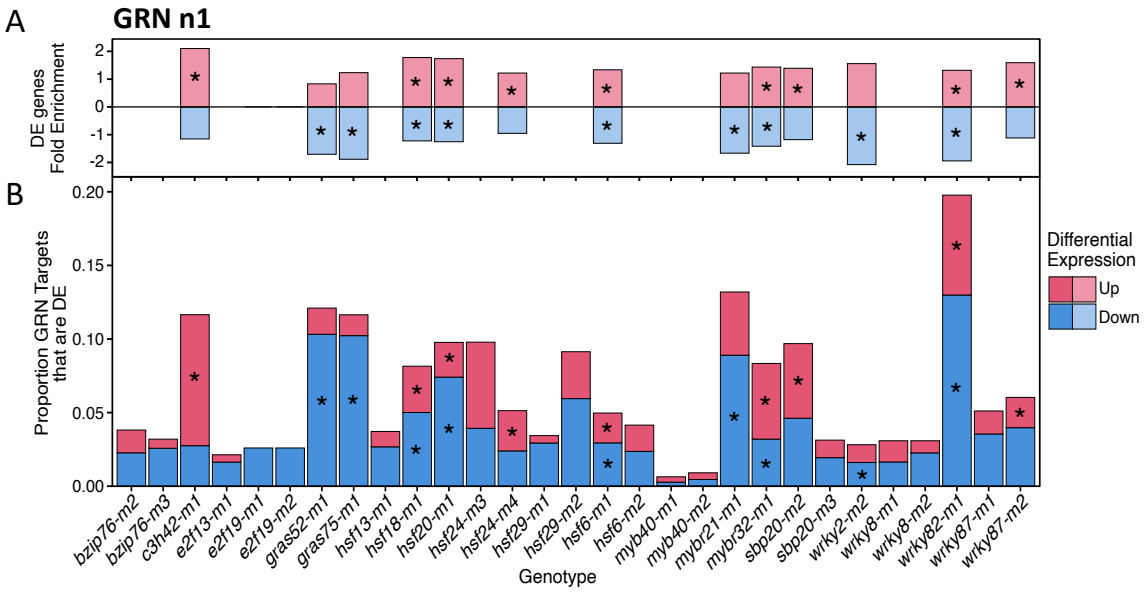


Figure 6. Enrichments of co-expression based GRN (n1) predictions for some TF mutants. The GRN predicted targets for each mutant were identified based on predicted GRN interactions being detected in at least one (n1) of 45 co-expression networks (Table S6). **(A)** For each mutant allele we calculated the fold-enrichment for predicted targets (the observed number of GRN predicted target genes that were significantly differentially expressed divided by the expected number of DE predicted targets) and the enrichments are shown for any mutants that exhibit significant enrichment (*: $p<0.05$). **(B)** The proportion of co-expression based GRN predicted targets that are up- (red) or down- (blue) regulated in each mutant is shown. For each allele, significant hypergeometric enrichment for up-regulated (red) and/or down-regulated (blue) DE target genes were marked (*).

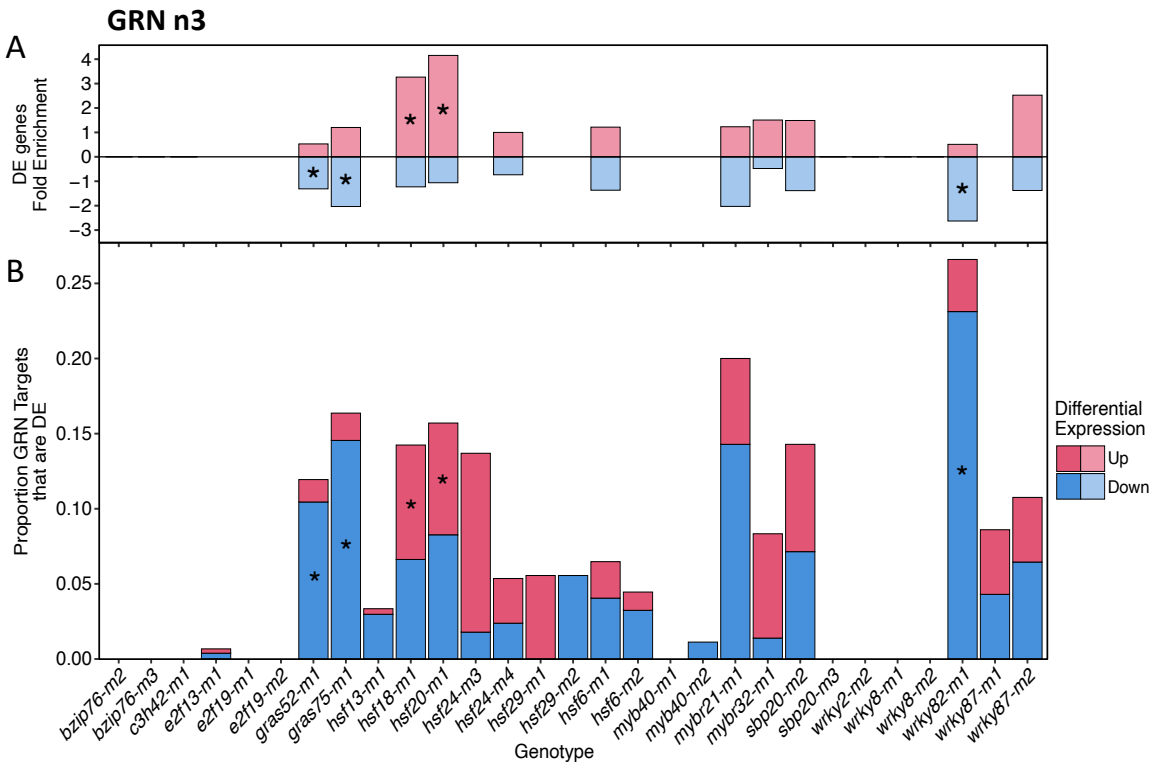


Figure S8. Enrichments of co-expression based GRN (n3) predictions for some TF mutants. The GRN predicted targets for each mutant were identified based predicted GRN interactions being detected in at least three (n3) of the 45 co-expression networks (Table S6). **(A)** For each mutant allele we calculated the fold-enrichment for predicted targets (the observed number of GRN predicted target genes that were significantly differentially expressed divided by the expected number of DE predicted targets). Enrichments are shown for any mutants that exhibit significant enrichment (*: $p < 0.05$) in at least 1/45 co-expression networks (n1), see Figure 6A. **(B)** The proportion of co-expression based GRN predicted targets that are up- (red) or down- (blue) regulated in each mutant is shown. For each allele, significant hypergeometric enrichment for up-regulated (red) and/or down-regulated (blue) DE target genes were marked (*).

Table S1. Maize TF mutant alleles isolated from the UniformMu population and transcriptome data obtained to test GRN predictions.

Transcription Factors				UniformMu Mutant Insertion Maize Genomic Data				Mutant RNAseq Sampling Information		
TF_Name	B73v4_gid	W22_gid	GRN Method	Mutant Allele	W22_MuID	B73v4_MuID	Stock	Pedigree	Tissue	Illumina
BAF60.21	Zm00001d015127	Zm00004b012791	yeast-one-hybrid	*baf60.21-m1*	chr5_77050725	mu1034781	UFMu-03236	BC2S2	tassel_stem	PE 150bp
BAF60.21	Zm00001d015127	Zm00004b012791	yeast-one-hybrid	*baf60.21-m2*	chr5_77050622	mu1092086	UFMu-11153	BC2S2	tassel_stem	PE 150bp
BZIP76	Zm00001d036736	Zm00004b029476	yeast-one-hybrid	*bzip76-m2*	chr6_95884230	mu1083967	UFMu-10010	BC2S2	embryo_imbibed	PE 150bp
BZIP76	Zm00001d036736	Zm00004b029476	yeast-one-hybrid	*bzip76-m3*	chr6_95884945	mu1081677	UFMu-09463	BC2S2	embryo_imbibed	PE 150bp
C3H42	Zm00001d008356	Zm00004b024707	yeast-one-hybrid	*c3h42-m1*	chr8_5818897	mu1068676	UFMu-08608	BC2S2	tassel_stem	PE 150bp
E2F19	Zm00001d027709	Zm00004b000391	yeast-one-hybrid	*e2f19-m1*	chr1_12237717	mu1040458	UFMu-04300	BC2S2	seedling_leaf	PE 150bp
E2F19	Zm00001d027709	Zm00004b000391	yeast-one-hybrid	*e2f19-m2*	chr1_12237755	mu1080409	UFMu-09504	BC2S2	seedling_leaf	PE 150bp
HSF24	Zm00001d032923	Zm00004b004268	yeast-one-hybrid	*hsf24-m3*	chr1_245656248	mu1065831	UFMu-08727	BC2S2	tassel	SE 50bp
HSF24	Zm00001d032923	Zm00004b004268	yeast-one-hybrid	*hsf24-m4*	chr1_245656475	mu1037205	UFMu-03655	BC2S2	tassel	SE 50bp
JMJ13	Zm00001d052933	Zm00004b023587	yeast-one-hybrid	*jmj13-m4*	chr4_210612840	mu1081210	UFMu-09466	BC2S2	tassel_stem	PE 150bp
MYB40	Zm00001d040621	Zm00004b016719	yeast-one-hybrid	*myb40-m1*	chr3_55921183	mu1090929	UFMu-11189	BC2S2	coleoptile_tip	PE 150bp
MYB40	Zm00001d040621	Zm00004b016719	yeast-one-hybrid	*myb40-m2*	chr3_55920569	mu1043293	UFMu-04991	BC2S2	coleoptile_tip	PE 150bp
E2F13	Zm00001d052288	Zm00004b023063	co-expression	*e2f13-m1*	chr4_192230417	mu1086360	UFMu-11121	S3	coleoptile_tip	PE 150bp
GRAS52	Zm00001d002573	Zm00004b006535	co-expression	*gras52-m1*	chr2_15799332	mu1037818	UFMu-03743	S3	embryo_imbibed	PE 150bp
GRAS75	Zm00001d006701	Zm00004b009784	co-expression	*gras75-m1*	chr2_213696662	mu1080849	UFMu-09435	S2	embryo_imbibed	PE 150bp
HSF13	Zm00001d027757	Zm00004b000433	co-expression	*hsf13-m1*	chr1_13648451	mu1085337	UFMu-10587	S3	seedling_leaf	PE 150bp
HSF18	Zm00001d016255	Zm00004b013608	co-expression	*hsf18-m1*	chr5_153354999	mu1086526	UFMu-10749	S3	embryo_imbibed	PE 150bp
HSF20	Zm00001d026094	Zm00004b040094	co-expression	*hsf20-m1*	chr10_136927897	mu1085720	UFMu-10752	S3	embryo_imbibed	PE 150bp
HSF29	Zm00001d016520	Zm00004b013825	co-expression	*hsf29-m1*	chr5_165555661	mu1023451	UFMu-02314	S3	embryo_imbibed	PE 150bp
HSF29	Zm00001d016520	Zm00004b013825	co-expression	*hsf29-m2*	chr5_165555817	mu1083642	UFMu-10398	S3	embryo_imbibed	PE 150bp
HSF6	Zm00001d016674	Zm00004b013941	co-expression	*hsf6-m1*	chr5_171668444	mu1048425	UFMu-06347	S3	embryo_imbibed	PE 150bp
HSF6	Zm00001d016674	Zm00004b013941	co-expression	*hsf6-m2*	chr5_171667932	mu1056797	UFMu-07611	S2	embryo_imbibed	PE 150bp
MYBR21	Zm00001d008602	Zm00004b024904	co-expression	*mybr21-m1*	chr8_14018025	mu1042351	UFMu-04838	S3	embryo_imbibed	PE 150bp
MYBR32	Zm00001d029963	Zm00004b002134	co-expression	*mybr32-m1*	chr1_96615987	mu1076168	UFMu-09083	S2	seedling_leaf	PE 150bp
SBP20	Zm00001d053890	Zm00004b024383	co-expression	*sbp20-m2*	chr4_249933323	mu1086430	UFMu-10892	S3	embryo_imbibed	PE 150bp
SBP20	Zm00001d053890	Zm00004b024383	co-expression	*sbp20-m3*	chr4_249932948	mu1091327	UFMu-11256	S2	embryo_imbibed	PE 150bp
WRKY2	Zm00001d016052	Zm00004b013450	co-expression	*wrky2-m2*	chr5_141290461	mu1025953	UFMu-11813	S3	coleoptile_tip	PE 150bp
WRKY8	Zm00001d053369	Zm00004b023980	co-expression	*wrky8-m1*	chr4_235052564	mu1048159	UFMu-06456	S3	embryo_imbibed	PE 150bp
WRKY8	Zm00001d053369	Zm00004b023980	co-expression	*wrky8-m2*	chr4_235053087	mu1077370	UFMu-08953	S3	embryo_imbibed	PE 150bp
WRKY82	Zm00001d038843	Zm00004b031112	co-expression	*wrky82-m1*	chr6_163095668	mu1081611	UFMu-09469	S3	embryo_imbibed	PE 150bp
WRKY87	Zm00001d052847	Zm00004b023521	co-expression	*wrky87-m1*	chr4_207897892	mu1067257	UFMu-08437	S3	embryo_imbibed	PE 150bp
WRKY87	Zm00001d052847	Zm00004b023521	co-expression	*wrky87-m2*	chr4_207898459	mu1091217	UFMu-12044	S3	embryo_imbibed	PE 150bp

Table S2. Gene expression values for the 22 TFs in different tissues. The expression value (CPM) for each of the 22 TFs was assessed based on prior sampling of tissues or developmental stages in B73 (Zhou et al. 2019). Values highlighted indicate the predicted expression level of TF genes in tissues sampled for RNA-seq in this study.

		24 h	6 DAS		seedling 11 DAS			v12					v14	0 DAP							14 DAP		27 DAP	
TF	RNAseq Tissue	embryo imbibed	coleoptile tip	radicle root	leaf	root	meristem	blade leaf	sheath	internode	tassel	auricle	ear	silk	spikelet	husk	tassel stem	floret	flag leaf	root	kernel	endosperm	embryo	endosperm
WRKY2	coleoptile_tip	3	11.94	6.24	1.44	3.73	6.24	0.61	0.93	1.69	2.45	1.18	1.46	1.4	2.1	1.31	2.56	0.87	1.16	0.3	3.08	2.4	3.86	1.27
MYB40	coleoptile_tip	4.88	45.73	21.18	7.93	1.56	19.14	5.71	4.08	0.78	14.11	12.82	10.39	6.77	9.75	4	1.25	8.52	59.8	2.03	22.29	23.86	5.51	26.23
E2F13	coleoptile_tip	0.07	20.7	9.56	0	6.1	34.15	0	0.03	0	5.91	0	34.84	0.14	2.92	0.08	0.08	0.05	0.05	2.26	1.1	0.48	6.5	0
GRAS52	embryo_imbibed	114.8	5.15	5.44	23.65	13.65	10.53	8.99	17.91	16.21	12.3	20.45	9.74	3.87	12.92	13.76	28.83	12.44	23.6	10.49	8.57	11.79	4.96	48.54
GRAS75	embryo_imbibed	163.04	32.39	21.85	17.72	7.29	14.77	32.9	26.97	9.64	7.27	18.24	18.26	12.51	16.36	16.08	60.99	19.78	71.3	4.41	47.58	62.12	20.78	58.19
MYBR21	embryo_imbibed	183.32	16.92	28.72	66.4	101	20.55	28.1	42.53	55.66	15.2	31.49	23.52	29.39	27.31	31.04	71.94	35.69	40.9	26.29	21.07	20.57	4.79	29.69
HSF18	embryo_imbibed	74.16	10.15	2.53	22.03	34.67	7.03	0.84	5.42	6.12	0.21	3.77	0.09	0.68	0.25	0.18	0.42	0.5	1.32	63.72	2.32	4.39	8.43	7.28
HSF29	embryo_imbibed	15.12	15.19	15.74	15.1	20.94	11.25	23.2	26.48	40.15	22.25	25.83	13.16	30.18	24.48	32.06	29.08	17.73	13.2	21.11	14.48	8.23	14.05	9.83
HSF6	embryo_imbibed	175.66	0.77	5.34	1.69	1.11	0.48	32.4	21.32	10.25	5.3	11.73	4.7	11.04	6.18	7.9	16.82	7.79	45	11.95	3.88	2.4	16.09	3
HSF20	embryo_imbibed	14.03	1.05	0.42	15.66	7.68	9.74	28.2	11.86	9.61	9	13.63	4.67	4.77	5.32	4.73	3.95	5.56	23.4	2.43	4.31	12.41	3.03	10.02
WRKY82	embryo_imbibed	43.53	6.97	15.77	30.95	16.55	1.51	7.73	12.26	18.3	1.52	11.92	0.53	1.4	2.67	8.87	14.29	5.29	10.8	43.58	4.05	0.14	0.22	0.64
WRKY87	embryo_imbibed	40.11	0.37	0.06	12.04	0.52	0.31	9.46	1.98	0.43	0.16	1.37	0.28	0	0.09	1.62	1.14	1.64	14.1	0.2	0.25	0.07	0.61	0.09
WRKY8	embryo_imbibed	6.21	3.9	3.13	0.87	4.71	4.63	0.89	2.33	2.58	1.89	3.39	2.58	2.62	1.7	1.62	3.11	1.23	2.73	1.07	1.27	0.96	4.74	0.64
SBP20	embryo_imbibed	37.53	11.77	13.47	5.3	11.04	23.5	1.4	10.37	18.71	16.9	7.5	11.79	2.19	8.48	4.81	16.96	7.93	1.67	37.25	3.72	2.33	6.12	0.82
BZIP76	embryo_imbibed	10.75	2.67	5.21	13.04	2.56	1.97	2.25	1.46	6.68	2.9	0.59	3.14	13.84	5	1.41	6.95	4.33	5.26	5.75	2.91	2.19	1.38	1.64
E2F19	seedling_leaf	15.8	9.67	10.3	16.98	9.87	15.8	13.5	13.46	11.65	12.62	10.99	23.11	14.59	14.61	9.99	15.21	8.8	22	11.43	14.73	14.12	13.01	19.94
HSF13	seedling_leaf	12.43	2.56	5.21	454.1	11.41	1.65	419	374.05	9.69	1.09	111.29	1.62	21.18	1.15	99.27	73.61	46.12	651	11.21	1.01	0	0.22	0.27
MYBR32	seedling_leaf	25.36	22.35	40.46	299.6	136.9	29.35	35.4	158.2	486.15	25.55	189.37	71.82	208.4	55.43	199.43	233.8	114.5	89.7	263.6	20.56	3.09	37.36	5.28
HSF24	tassel	61.01	32.33	39.41	76.39	52.68	120.25	457	250.25	292.19	96.74	188.37	80.66	82.65	100.08	73.67	109.6	88.97	400	155.9	167.1	208.16	265.23	121.11
C3H42	tassel_stem	3.9	7.39	10.4	2.18	4.56	9.54	3.46	5.16	27.27	7.45	3.89	6.87	8.24	7.17	6.72	9.59	2.83	4.7	7.76	9.12	11.18	4.9	6.47
BAF60.21	tassel_stem	14.59	22.83	20.86	4.24	12.8	40.22	14.1	14.71	8.16	11.74	14.07	19.75	14.05	10.77	13.31	16.32	11.26	24.1	14.86	7.51	6.1	16.7	3.92
JMJ13	tassel_stem	27.45	19.45	14.87	0.87	11.5	21.1	6.13	8.24	6.9	7.03	7.03	18.32	10.68	11.42	6.64	6.81	6.47	7.79	8.06	20.09	23.52	13.94	16.03

Table S3. Transcription factor differential gene expression data. Differential expression data averaged across biological replicates (N-rep) in mutant allele (CPM) and W22 control (W22_CPM) samples (log2fc: log2 fold change of mutant to control, lfcSE: log fold change standard error and FDR adjusted p-value included).

		Transcript Abundance							
		mutant		W22		RNAseq DE Data			
Genotype	Tissue	CPM	N	CPM	N	log2fc	lfcSE	pvalue	DE
baf60.21-m1	tassel_stem	37.49	3	26.96	3	0.480	0.114	8.22E-04	Not_DE
baf60.21-m2	tassel_stem	15.22	3	26.96	3	-0.829	0.118	1.16E-09	Not_DE
bzip76-m2	leaf	7.92	3	7.41	4	0.064	0.217	0.90064	Not_DE
bzip76-m3	leaf	3.80	3	7.41	4	-0.954	0.237	0.0011	Not_DE
c3h42-m1	tassel_stem	5.05	3	3.73	3	0.443	0.173	0.06283	Not_DE
e2f13-m1	coleoptile_tip	5.77	3	8.90	4	-0.607	0.159	0.00362	Not_DE
e2f19-m1	leaf	4.13	3	9.22	4	-1.125	0.217	8.47E-06	DE_Down
e2f19-m2	leaf	6.82	3	9.22	4	-0.471	0.203	0.11291	Not_DE
gras52-m1	embryo_imbibed	0.72	2	12.33	4	-4.072	0.355	4.83E-28	DE_Down
gras75-m1	embryo_imbibed	54.10	3	130.48	4	-1.266	0.142	4.06E-17	DE_Down
hsf13-m1	leaf	45.52	3	92.53	4	-0.989	0.110	4.03E-17	Not_DE
hsf18-m1	embryo_imbibed	7.69	3	2.83	4	1.373	0.290	2.90E-05	DE_Up
hsf20-m1	embryo_imbibed	0.89	3	0.61	4	0.514	0.692	0.63301	Not_DE
hsf24-m3	tassel	51.22	3	41.15	3	0.045	0.168	0.87331	Not_DE
hsf24-m4	tassel	21.24	3	41.15	3	-1.104	0.172	9.37E-09	DE_Down
hsf29-m1	embryo_imbibed	32.75	3	27.83	4	0.201	0.168	0.44145	Not_DE
hsf29-m2	embryo_imbibed	27.94	3	27.83	4	-0.025	0.169	0.93601	Not_DE
hsf6-m1	embryo_imbibed	0.04	3	0.48	4	-3.748	1.245	0.01579	DE_Down
hsf6-m2	embryo_imbibed	0.34	3	0.48	4	-0.479	0.772	0.72818	Not_DE
jmj13-m4	tassel_stem	8.90	3	14.14	3	-0.669	0.133	1.13E-04	Not_DE
myb40-m1	coleoptile_tip	13.17	3	11.57	4	0.184	0.147	0.89794	Not_DE
myb40-m2	coleoptile_tip	8.76	3	11.57	4	-0.372	0.151	0.34221	Not_DE
mybr21-m1	embryo_imbibed	14.29	3	60.91	4	-2.075	0.129	1.64E-55	DE_Down
mybr32-m1	leaf	61.19	3	141.69	4	-1.187	0.083	2.06E-43	DE_Down
sbp20-m2	embryo_imbibed	9.39	3	14.03	4	-0.620	0.160	8.08E-04	Not_DE
sbp20-m3	embryo_imbibed	14.90	3	14.03	4	0.065	0.160	0.82448	Not_DE
wrky2-m2	coleoptile_tip	2.21	3	2.29	4	-0.059	0.258	0.98609	Not_DE
wrky8-m1	embryo_imbibed	19.28	3	7.15	4	1.452	0.295	2.40E-05	DE_Up
wrky8-m2	embryo_imbibed	5.38	3	7.15	4	-0.330	0.301	0.51222	Not_DE
wrky82-m1	embryo_imbibed	7.19	3	34.36	4	-2.297	0.262	8.04E-17	DE_Down
wrky87-m1	embryo_imbibed	8.83	3	5.99	4	0.535	0.252	0.10716	Not_DE
wrky87-m2	embryo_imbibed	8.84	3	5.99	4	0.571	0.252	0.07387	Not_DE

Table S4. Mutant allele transcriptome assembly predicted transcript structure. Transcripts are described either 5' or 3' of *Mu*.

	Mutant Transcript Structure	
Allele	**Assembled transcripts matching 5' of *Mu* insertion**	**Assembled transcripts matching 3' of *Mu* insertion**
baf60.21-m1	No assembled transcript detected	Assembled transcript initiates within Mu sequence and contains full CDS: predicted ORF is out-of-frame
baf60.21-m2	No assembled transcript detected	Assembled transcript initiates within Mu sequence and contains partial CDS: predicted ORF is in-frame
bzip76-m2	No assembled transcript detected	Assembled transcript initiates within Mu sequence and contains partial CDS: predicted ORF is in-frame
bzip76-m3	No assembled transcript detected	Assembled transcript initaties in variable unmapped sequence without detection of Mu and contains partial CDS: predicted ORF is in-frame
c3h42-m1	Alternative splicing. Two assembled transcripts initiate from the gene TSS 5' of the intronic insertion. Transcript 1 contains variable additional sequence near the insertion site and reads through to the 3' gene end and transcript 2 produces a truncated transcript with a premature stop within Mu. Both transcripts predicted ORFs are in-frame.	Alternative splicing. Four assembled transcripts initiate 3' of the intronic insertion. 3/4 transcripts initiate within Mu and read through to the 3' gene end: 1/3 transcripts predicted ORF is in-frame, 2/3 transcripts predicted ORF is out-of-frame. Transcript 4 produces a truncated transcript initiating after Mu: predicted ORF is out-of-frame.
e2f19-m1	No assembled transcript detected	Assembled transcript initiates within partial CDS and reads through to the 3' gene end, no 5'UTR detected: predicted ORF is in-frame
e2f19-m2	No assembled transcript detected	Assembled transcript initiates within partial CDS and reads through to the 3' gene end, no 5'UTR detected: predicted ORF is in-frame
hsf24-m3	Assembled transcript initiates in the 5'UTR and prematurely stop within Mu: predicted ORF in-frame	Assembled transcript initaties within partial CDS and reads through to the 3' gene end: predicted ORF is in-frame
hsf24-m4	No assembled transcript detected	Assembled transcript initiates within partial CDS and reads through to the 3' gene end, no 5'UTR detected: predicted ORF is in-frame
jmj13-m4	Assembled transcript initiates in the 5'UTR and prematurely stops within Mu: predicted ORF in-frame	Assembled transcript initiates within Mu sequence and contains partial CDS: predicted ORF is out-of-frame
myb40-m1	Assembled transcript initiates in the 5'UTR and prematurely stops within Mu: predicted ORF in-frame	Assembled transcript initiates within Mu sequence and contains partial CDS: predicted ORF is out-of-frame.
myb40-m2	No assembled transcript detected	Assembled transcript initiates within partial CDS and reads through to the 3' gene end, no 5'UTR detected: predicted ORF is in-frame
e2f13-m1	Assembled transcript initiates in partial CDS and prematurely stops within Mu: predicted ORF in-frame	Assembled transcript initiates within Mu sequence and reads through to the 3' gene end: predicted ORF is out-of-frame
gras52-m1	Assembled transcript initiates from the gene TSS and prematurely stop within Mu: predicted ORF in-frame	Assembled transcript initaties within partial CDS and reads through to the 3' gene end: predicted ORF is in-frame
gras75-m1	Assembled transcript initiates from the gene TSS and prematurely stops within Mu: predicted ORF in-frame	No assembled transcript detected
hsf13-m1	Assembled transcript initiates in partial CDS and prematurely stops within Mu: predicted ORF in-frame	Assembled transcript initiates within Mu sequence and reads through to the 3' gene end: predicted ORF is in-frame
hsf18-m1	Assembled transcript initiates in the 5'UTR and prematurely stops near Mu insertion site, no Mu sequence detected: predicted ORF in-frame	Assembled transcript initiates within Mu sequence and reads through to the 3' gene end: predicted ORF is in-frame
hsf20-m1	Assembled transcript initiates in the 5'UTR and prematurely stops within Mu: predicted ORF in-frame	Assembled transcript initaties within partial CDS and reads through to the 3' gene end: predicted ORF is out-of-frame
hsf29-m1	No assembled transcript detected	Assembled transcript initiates within Mu sequence and contains partial 5'UTR and full CDS sequence: predicted ORF is out-of-frame
hsf29-m2	No assembled transcript detected	Assembled transcript initaties in variable unmapped sequence without detection of Mu and contains partial CDS: predicted ORF is out-of-frame
hsf6-m1	No assembled transcript detected- coverage too low in the wild-type and mutant to assemble transcripts	No assembled transcript detected- coverage too low in the wild-type and mutant to assemble transcripts
hsf6-m2	No assembled transcript detected- coverage too low in the wild-type and mutant to assemble transcripts	No assembled transcript detected- coverage too low in the wild-type and mutant to assemble transcripts
mybr21-m1	Assembled transcript initiates in the 5'UTR and prematurely stop within Mu: predicted ORF in-frame	Assembled transcript initaties within partial CDS and reads through to the 3' gene end: predicted ORF is out-of-frame
mybr32-m1	Assembled transcript initiates from the gene TSS and prematurely stops within Mu: predicted ORF in-frame	Assembled transcript initiates within Mu sequence and contains partial CDS: predicted ORF is out-of-frame
sbp20-m2	Assembled transcript initiates from the gene TSS and prematurely stops within Mu: predicted ORF in-frame	Assembled transcript initiates within Mu sequence and contains partial CDS: predicted ORF is out-of-frame
sbp20-m3	Assembled transcript initiates from the gene TSS and prematurely stops within Mu: predicted ORF in-frame	Assembled transcript initiates within Mu sequence and contains partial CDS: predicted ORF is out-of-frame
wrky2-m2	Assembled transcript initiates from the gene TSS and prematurely stops near Mu insertion site, no Mu sequence detected: predicted ORF in-frame	Assembled transcript initiates within Mu sequence and reads through to the 3' gene end: predicted ORF is out-of-frame
wrky8-m1	Assembled transcript initiates from the gene TSS and prematurely stops within Mu: predicted ORF in-frame	Assembled transcript initiates within Mu sequence and reads through to the 3' gene end: predicted ORF is out-of-frame
wrky8-m2	Assembled transcript initiates from the gene TSS and prematurely stops with variable unmapped sequence near the Mu insertion site, no Mu sequence detected: predicted ORF in-frame	Assembled transcript initiates within Mu sequence and reads through to the 3' gene end: predicted ORF is out-of-frame
wrky82-m1	Assembled transcript initiates from the gene TSS and prematurely stops within Mu: predicted ORF in-frame	No assembled transcript detected
wrky87-m1	Assembled transcript initiates from the gene TSS and prematurely stops within Mu: predicted ORF in-frame	Assembled transcript initiates within Mu sequence and reads through to the 3' gene end: predicted ORF is out-of-frame
wrky87-m2	Assembled transcript initiates from the gene TSS and prematurely stops within Mu: predicted ORF in-frame	Assembled transcript initiates within Mu sequence and reads through to the 3' gene end: predicted ORF is out-of-frame

* Putative ORF is based on the first start codon in the assembled transcript

CONCLUSION

The initial goal of my doctoral research was to experimentally validate GRN predictions in maize by perturbing the networks *in-vivo* (Chapter III). I sought to isolate loss-of-function mutants for predicted TF regulators to test expression of predicted targets *in-vivo*. Results from this study would help determine the relative value of GRN predictions from various methods and identify biologically relevant TF-target gene interactions in maize. Putative loss-of-function mutants were isolated from the UniformMu transposon-indexed population for many maize transcription factor genes. Significant time and effort went into identifying homozygous transmissible alleles and selecting lines with a reduced mutational load using genetics and molecular genotyping. We successfully obtained a collection of putative loss-of-function mutants for maize transcription factors with GRN predicted targets. For all the mutants we isolated we were hoping to characterize TF function by observing visible phenotypes that co-segregated with the mutation. Unfortunately, none of our mutants exhibited a visible phenotype that segregated with the *Mu*-induced mutation in the TF gene. It would have been interesting to learn more about a TFs regulatory role by linking target genes to the phenotype observed.

Although none of these TF mutants exhibited a visible phenotype, it was still possible to assess whether these mutants resulted in loss-of TF function and test expression of GRN predicted target genes. We attempted to determine if these homozygous *Mu*-induced mutations resulted in loss-of-function mutations by performing RT-PCR with primers flanking the *Mu* insertion. We had limited success in determining how *Mu* was affecting transcript structure and abundance with this method. We initially thought we may recover mutant transcripts by RT-PCR with primers flanking the *Mu* if there was read-through of partial *Mu* sequence, but in most cases, we failed to recover amplified sequences using this approach. We decided to move forward with RNA-seq of these mutants to determine the transcript produced by the mutant allele and to test expression of predicted targets. If the GRN predicted targets were directly bound and regulated by the predicted TF regulators we expected target genes to exhibit differential expression (DE) in the absence of the TF. After identifying genes that were DE between TF mutant and wild-type, we found minimal overlap of DE genes and predicted GRN targets. The results of my initial thesis project provide some insights into GRN network predictions but did not provide substantial advancements towards our understanding of the specific

functions of the TFs that were tested. We reasoned that experimental validation of GRN predictions might require higher order TF mutants to account for genetic redundancy in maize.

While performing this study, we generated a collection of loss-of-function mutants in several TF genes in maize, but the direction of the remainder of my doctoral research was unclear. However, upon further analysis of our existing RNA-seq data we identified variable mapping of reads that flanked the *Mu* insertion site. To better understand the structure of the transcripts produced by the mutant alleles we performed a transcriptome assembly of RNA-seq reads for each mutant allele. We found assembled transcripts of many genes with these *Mu*-induced mutations aligned to either gene sequence 5' or 3' of the annotated *Mu* insertion, but not both. There was evidence that many of these mutants produced these split transcripts flanking the *Mu*-indexed insertion site. Next, we determined if these transcripts contained *Mu* sequence and found evidence for transcripts initiating and terminating within *Mu* sequence. We found that transcripts initiating within *Mu* sequence, likely directed from an outward-reading promoter within the *Mu* element, was a reproducible finding for mutants in the UniformMu population. This ultimately led to having the genetic materials to study how this outward-reading promoter in *Mu* could affect gene expression and influence how *Mu* preferentially inserts in or near genes (Chapter II).

To determine if this *Mu* promoter could direct expression levels and tissue-specific expression patterns that were different than that of the normal gene promoter, we compared expression levels and patterns of the gene transcripts in both the mutant and wild-type. We found that the mutant partial gene promoter transcripts with termination in *Mu* exhibited very similar expression levels and tissue-specific patterns to wild-type, which suggests that the normal gene promoter in the mutants could direct transcription similar to wild-type and there were minimal effects from transcript turnover (i.e., nonsense mediated decay). We also found that mutant transcripts produced from the *Mu* promoter had similar to wild-type abundances and patterns of expression. Previous work had hypothesized that this promoter may be more of a minimal promoter that can interact with gene regulatory sequences. In our study, we find evidence that this hypothesis may be true. Results from our study indicate that *Mu* may be able to mimic

normal gene expression by direct transcription of adjacent gene sequence from an active outward-reading promoter. These findings have implications for future transposon biology. The behavior of *Mu* provides a system to study how regulatory sequences within transposons are co-opted to regulate gene expression, which can provide a means for transposon survival in the host genome by masking the effects of their insertion. Further analyses will lead to a greater understanding of how these regulatory elements in TEs may contribute to gene regulation at a genome-wide level.

BIBLIOGRAPHY

1. Spitz, F. & Furlong, E. E. M. Transcription factors: from enhancer binding to developmental control. *Nat. Rev. Genet.* **13**, 613–626 (2012).
2. Fedoroff, N. V. Presidential address. Transposable elements, epigenetics, and genome evolution. *Science* **338**, 758–767 (2012).
3. Kidwell, M. G. & Lisch, D. Transposable elements as sources of variation in animals and plants. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 7704–7711 (1997).
4. Bourque, G. *et al.* Ten things you should know about transposable elements. *Genome Biol.* **19**, 199 (2018).
5. Chuong, E. B., Elde, N. C. & Feschotte, C. Regulatory activities of transposable elements: from conflicts to benefits. *Nat. Rev. Genet.* **18**, 71–86 (2016).
6. Sundaram, V. & Wysocka, J. Transposable elements as a potent source of diverse cis-regulatory sequences in mammalian genomes. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **375**, 20190347 (2020).
7. Fueyo, R., Judd, J., Feschotte, C. & Wysocka, J. Roles of transposable elements in the regulation of mammalian transcription. *Nat. Rev. Mol. Cell Biol.* (2022) doi:10.1038/s41580-022-00457-y.
8. Wells, J. N. & Feschotte, C. A Field Guide to Eukaryotic Transposable Elements. *Annu. Rev. Genet.* **54**, 539–561 (2020).
9. Østergaard, L. & Yanofsky, M. F. Establishing gene function by mutagenesis in Arabidopsis thaliana. *Plant J.* **39**, 682–696 (2004).
10. Parinov, S. & Sundaresan, V. Functional genomics in Arabidopsis: large-scale insertional mutagenesis complements the genome sequencing project. *Curr. Opin. Biotechnol.* **11**, 157–161 (2000).
11. Brutnell, T. P. Transposon tagging in maize. *Funct. Integr. Genomics* **2**, 4–12 (2002).
12. Tadege, M., Ratet, P. & Mysore, K. S. Insertional mutagenesis: a Swiss Army knife for functional genomics of Medicago truncatula. *Trends Plant Sci.* **10**, 229–235 (2005).
13. Lisch, D. Mutator transposons. *Trends Plant Sci.* **7**, 498–504 (2002).
14. Lisch, D. & Jiang, N. Mutator and MULE Transposons. *Microbiol Spectr* **3**, MDNA3–0032–2014 (2015).
15. Lisch, D. Regulation of the Mutator System of Transposons in Maize. in *Plant Transposable Elements: Methods and Protocols* (ed. Peterson, T.) 123–142 (Humana Press, 2013).
16. Walbot, V. The Mutator Transposable Element Family of Maize. in *Genetic Engineering: Principles and Methods* (ed. Setlow, J. K.) 1–37 (Springer US, 1991).
17. Bennetzen, J. L. The Mutator transposable element system of maize. *Curr. Top. Microbiol. Immunol.* **204**, 195–229 (1996).
18. Chandler, V. L. & Hardeman, K. J. The Mu elements of Zea mays. *Adv. Genet.* **30**, 77–122 (1992).
19. Robertson, D. S. Characterization of a mutator system in maize. *Mutat. Res./Fundam. Mol. Mech. Mutag.* **51**, 21–28 (1978).
20. Schnable, P. S. & Peterson, P. A. The Mutator-Related Cy Transposable Element of Zea Mays L. Behaves as a near-Mendelian Factor. *Genetics* **120**, 587–596 (1988).
21. Cresse, A. D., Hulbert, S. H., Brown, W. E., Lucas, J. R. & Bennetzen, J. L. Mu1-related transposable elements of maize preferentially insert into low copy number DNA. *Genetics* **140**, 315–324 (1995).

22. McCarty, D. R. *et al.* Steady-state transposon mutagenesis in inbred maize: Maize steady-state transposon mutagenesis. *Plant J.* **44**, 52–61 (2005).
23. Marcon, C. *et al.* BonnMu: A Sequence-Indexed Resource of Transposon-Induced Maize Mutations for Functional Genomics Studies. *Plant Physiol.* **184**, 620–631 (2020).
24. Williams-Carrier, R. *et al.* Use of Illumina sequencing to identify transposon insertions underlying mutant phenotypes in high-copy Mutator lines of maize: Illumina-HTS identification of Mu tagged genes. *Plant J.* **152**, no–no (2010).
25. May, B. P. *et al.* Maize-targeted mutagenesis: A knockout resource for maize. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 11541–11546 (2003).
26. Fedoroff, N. V. & Chandler, V. Inactivation of Maize Transposable Elements. in *Homologous Recombination and Gene Silencing in Plants* (ed. Paszkowski, J.) 349–385 (Springer Netherlands, 1994).
27. Strommer, J. N., Hake, S., Bennetzen, J., Taylor, W. C. & Freeling, M. Regulatory mutants of the maize Adh1 gene caused by DNA insertions. *Nature* **300**, 542–544 (1982).
28. Bennetzen, J. L. Transposable element Mu1 is found in multiple copies only in Robertson's Mutator maize lines. *J. Mol. Appl. Genet.* **2**, 519–524 (1984).
29. Taylor, L. P. & Walbot, V. Isolation and characterization of a 1.7-kb transposable element from a mutator line of maize. *Genetics* **117**, 297–307 (1987).
30. Settles, A. M., Latshaw, S. & McCarty, D. R. Molecular analysis of high-copy insertion sites in maize. *Nucleic Acids Res.* **32**, e54 (2004).
31. Lisch, D. Mutator transposons. *Trends Plant Sci.* **7**, 498–504 (2002).
32. Fernandes, J. *et al.* Genome-wide mutagenesis of Zea mays L. using RescueMu transposons. *Genome Biol.* **5**, R82 (2004).
33. Alleman, M. & Freeling, M. The Mu transposable elements of maize: evidence for transposition and copy number regulation during development. *Genetics* **112**, 107–119 (1986).
34. Fedoroff, N. V. & Chandler, V. Inactivation of Maize Transposable Elements. in *Homologous Recombination and Gene Silencing in Plants* (ed. Paszkowski, J.) 349–385 (Springer Netherlands, 1994).
35. Talbert, L. E., Patterson, G. I. & Chandler, V. L. Mu transposable elements are structurally diverse and distributed throughout the genus Zea. *J. Mol. Evol.* **29**, 28–39 (1989).
36. Fleenor, D., Spell, M., Robertson, D. & Wessler, S. Nucleotide sequence of the maize Mutator element, Mu8. *Nucleic Acids Res.* **18**, 6725 (1990).
37. Tan, B.-C. *et al.* Identification of an active new mutator transposable element in maize. *G3* **1**, 293–302 (2011).
38. Dietrich, C. R. *et al.* Maize Mu transposons are targeted to the 5' untranslated region of the gl8 gene and sequences flanking Mu target-site duplications exhibit nonrandom nucleotide composition throughout the genome. *Genetics* **160**, 697–716 (2002).
39. Zhao, D., Ferguson, A. A. & Jiang, N. What makes up plant genomes: The vanishing line between transposable elements and genes. *Biochim. Biophys. Acta* **1859**, 366–380 (2016).
40. Hirsch, C. D. & Springer, N. M. Transposable element influences on gene expression in plants. *Biochim. Biophys. Acta* **1860**, 157–165 (2017).
41. Vollbrecht, E. *et al.* Genome-wide distribution of transposed Dissociation elements in maize. *Plant Cell* **22**, 1667–1685 (2010).

42. Tock, A. J. & Henderson, I. R. Hotspots for Initiation of Meiotic Recombination. *Front. Genet.* **9**, 521 (2018).
43. Underwood, C. J., Henderson, I. R. & Martienssen, R. A. Genetic and epigenetic variation of transposable elements in Arabidopsis. *Curr. Opin. Plant Biol.* **36**, 135–141 (2017).
44. Lisch, D., Chomet, P. & Freeling, M. Genetic characterization of the Mutator system in maize: behavior and regulation of Mu transposons in a minimal line. *Genetics* **139**, 1777–1796 (1995).
45. Hershberger, R. J. *et al.* Characterization of the major transcripts encoded by the regulatory MuDR transposable element of maize. *Genetics* **140**, 1087–1098 (1995).
46. Chomet, P., Lisch, D., Hardeman, K. J., Chandler, V. L. & Freeling, M. Identification of a regulatory transposon that controls the Mutator transposable element system in maize. *Genetics* **129**, 261–270 (1991).
47. Martienssen, R. & Baron, A. Coordinate suppression of mutations caused by Robertson's mutator transposons in maize. *Genetics* **136**, 1157–1170 (1994).
48. Martienssen, R. A., Barkan, A., Freeling, M. & Taylor, W. C. Molecular cloning of a maize gene involved in photosynthetic membrane organization that is regulated by Robertson's Mutator. *EMBO J.* **8**, 1633–1639 (1989).
49. Barkan, A. & Martienssen, R. A. Inactivation of Maize Transposon Mu Suppresses a Mutant Phenotype by Activating an Outward-Reading Promoter Near the End of Mu1. *Proceedings of the National Academy of Sciences* **88**, 3502–3506 (1991).
50. Martienssen, R., Barkan, A., Taylor, W. C. & Freeling, M. Somatically heritable switches in the DNA modification of Mu transposable elements monitored with a suppressible mutant in maize. *Genes Dev.* **4**, 331–343 (1990).
51. Chomet, P., Lisch, D., Hardeman, K. J., Chandler, V. L. & Freeling, M. Identification of a regulatory transposon that controls the Mutator transposable element system in maize. *Genetics* **129**, 261–270 (1991).
52. Pooma, W., Gersos, C. & Grotewold, E. Transposon insertions in the promoter of the Zea mays a1 gene differentially affect transcription by the Myb factors P and C1. *Genetics* **161**, 793–801 (2002).
53. Girard, L. & Freeling, M. Mutator-suppressible alleles of rough sheath1 and liguleless3 in maize reveal multiple mechanisms for suppression. *Genetics* **154**, 437–446 (2000).
54. Lowe, B., Mathern, J. & Hake, S. Active Mutator elements suppress the knotted phenotype and increase recombination at the Kn1-O tandem duplication. *Genetics* **132**, 813–822 (1992).
55. Cui, X. *et al.* Alternative transcription initiation sites and polyadenylation sites are recruited during Mu suppression at the rf2a locus of maize. *Genetics* **163**, 685–698 (2003).
56. Das, L. & Martienssen, R. Site-selected transposon mutagenesis at the hcf106 locus in maize. *Plant Cell* **7**, 287–294 (1995).
57. Settles, A. M., Baron, A., Barkan, A. & Martienssen, R. A. Duplication and suppression of chloroplast protein translocation genes in maize. *Genetics* **157**, 349–360 (2001).
58. Chandler, V. L. & Walbot, V. DNA modification of a maize transposable element correlates with loss of activity. *Proc. Natl. Acad. Sci. U. S. A.* **83**, 1767–1771 (1986).
59. Bennetzen, J. L. Covalent DNA modification and the regulation of Mutator element transposition in maize. *Mol. Gen. Genet.* **208**, 45–51 (1987).

60. Bennetzen, J. L., Brown, W. E. & Springer, P. S. The State of DNA Modification within and Flanking Maize Transposable Elements. in *Plant Transposable Elements* (eds. Nelson, O., Wilson, C. M. & Saslaw, C. G.) 237–250 (Springer US, 1988).
61. Slotkin, R. K., Freeling, M. & Lisch, D. Mu killer causes the heritable inactivation of the Mutator family of transposable elements in Zea mays. *Genetics* **165**, 781–797 (2003).
62. Slotkin, R. K., Freeling, M. & Lisch, D. Heritable transposon silencing initiated by a naturally occurring transposon inverted duplication. *Nat. Genet.* **37**, 641–644 (2005).
63. Hollister, J. D. & Gaut, B. S. Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression. *Genome Res.* **19**, 1419–1428 (2009).
64. Settles, A. M. *et al.* Sequence-indexed mutations in maize using the UniformMu transposon-tagging population. *BMC Genomics* **8**, 116 (2007).
65. Springer, N. M. *et al.* The maize W22 genome provides a foundation for functional genomics and transposon biology. *Nat. Genet.* (2018) doi:10.1038/s41588-018-0158-0.
66. McCarty, D. R. *et al.* Genetic and molecular analyses of UniformMu transposon insertion lines. *Methods Mol. Biol.* **1057**, 157–166 (2013).
67. McCarty, D. R., Liu, P. & Koch, K. E. The UniformMu Resource: Construction, Applications, and Opportunities. in *The Maize Genome* (eds. Bennetzen, J., Flint-Garcia, S., Hirsch, C. & Tuberosa, R.) 131–142 (Springer International Publishing, 2018).
68. McCarty, D. R. *et al.* Mu-seq: sequence-based mapping and identification of transposon induced mutations. *PLoS One* **8**, e77172 (2013).
69. May, B. P. *et al.* Maize-targeted mutagenesis: A knockout resource for maize. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 11541–11546 (2003).
70. Mejia-Guerra, M. K., Pomeranz, M., Morohashi, K. & Grotewold, E. From plant gene regulatory grids to network dynamics. *Biochim. Biophys. Acta* **1819**, 454–465 (2012).
71. Barabási, A.-L. & Oltvai, Z. N. Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.* **5**, 101–113 (2004).
72. Springer, N., de León, N. & Grotewold, E. Challenges of Translating Gene Regulatory Information into Agronomic Improvements. *Trends Plant Sci.* (2019) doi:10.1016/j.tplants.2019.07.004.
73. Morohashi, K., Xie, Z. & Grotewold, E. Gene-specific and genome-wide ChIP approaches to study plant transcriptional networks. *Methods Mol. Biol.* **553**, 3–12 (2009).
74. O'Malley, R. C. *et al.* Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape. *Cell* **165**, 1280–1292 (2016).
75. Haque, S., Ahmad, J. S., Clark, N. M., Williams, C. M. & Sozzani, R. Computational prediction of gene regulatory networks in plant growth and development. *Curr. Opin. Plant Biol.* **47**, 96–105 (2018).
76. Krouk, G., Lingeman, J., Colon, A. M., Coruzzi, G. & Shasha, D. Gene regulatory networks in plants: learning causality from time and perturbation. *Genome Biol.* **14**, 123 (2013).
77. Kemmeren, P. *et al.* Large-scale genetic perturbations reveal regulatory networks and an abundance of gene-specific repressors. *Cell* **157**, 740–752 (2014).

78. Thompson, D., Regev, A. & Roy, S. Comparative analysis of gene regulatory networks: from network reconstruction to evolution. *Annu. Rev. Cell Dev. Biol.* **31**, 399–428 (2015).
79. Scherens, B. & Goffeau, A. The uses of genome-wide yeast mutant collections. *Genome Biol.* **5**, 229 (2004).
80. Muhammad, D., Schmittling, S., Williams, C. & Long, T. A. More than meets the eye: Emergent properties of transcription factors networks in Arabidopsis. *Biochim. Biophys. Acta* **1860**, 64–74 (2017).
81. Schnable, J. C., Pedersen, B. S., Subramaniam, S. & Freeling, M. Dose-sensitivity, conserved non-coding sequences, and duplicate gene retention through multiple tetraploidies in the grasses. *Front. Plant Sci.* **2**, 2 (2011).
82. Tu, X. *et al.* Reconstructing the maize leaf regulatory network using ChIP-seq data of 104 transcription factors. *Nat. Commun.* **11**, 5089 (2020).
83. Swanson-Wagner, R. A. *et al.* Pervasive gene content variation and copy number variation in maize and its undomesticated progenitor. *Genome Res.* **20**, 1689–1699 (2010).
84. Wu, W.-S. & Lai, F.-J. Functional redundancy of transcription factors explains why most binding targets of a transcription factor are not affected when the transcription factor is knocked out. *BMC Syst. Biol.* **9 Suppl 6**, S2 (2015).
85. Bessereau, J.-L. Insertional mutagenesis in C. elegans using the Drosophila transposon Mos1: a method for the rapid identification of mutated genes. *Methods Mol. Biol.* **351**, 59–73 (2006).
86. Bessereau, J. L. *et al.* Mobilization of a Drosophila transposon in the Caenorhabditis elegans germ line. *Nature* **413**, 70–74 (2001).
87. Thibault, S. T. *et al.* A complementary transposon tool kit for Drosophila melanogaster using P and piggyBac. *Nat. Genet.* **36**, 283–287 (2004).
88. Cooley, L., Kelley, R. & Spradling, A. Insertional mutagenesis of the Drosophila genome with single P elements. *Science* **239**, 1121–1128 (1988).
89. Cain, A. K. *et al.* A decade of advances in transposon-insertion sequencing. *Nat. Rev. Genet.* **21**, 526–540 (2020).
90. Guo, Y. *et al.* Integration profiling of gene function with dense maps of transposon integration. *Genetics* **195**, 599–609 (2013).
91. Michel, A. H. *et al.* Functional mapping of yeast genomes by saturated transposition. *Elife* **6**, (2017).
92. Ivics, Z. *et al.* Transposon-mediated genome manipulation in vertebrates. *Nat. Methods* **6**, 415–422 (2009).
93. Brutnell, T. P. & Conrad, L. J. Transposon tagging using Activator (Ac) in maize. *Methods Mol. Biol.* **236**, 157–176 (2003).
94. Bensen, R. J. *et al.* Cloning and characterization of the maize An1 gene. *Plant Cell* **7**, 75–84 (1995).
95. Briggs, S. & Meeley, B. Reverse genetics for maize. *Maize Genetics Cooperation Newsletter* vol. 69 67–82 (1995).
96. Bennetzen, J. L. & Springer, P. S. The generation of Mutator transposable element subfamilies in maize. *Theor. Appl. Genet.* **87**, 657–667 (1994).
97. Rabinowicz, P. D. *et al.* Differential methylation of genes and retrotransposons facilitates shotgun sequencing of the maize genome. *Nat. Genet.* **23**, 305–308 (1999).
98. Hanley, S. *et al.* Identification of transposon-tagged genes by the random sequencing of Mutator-tagged DNA fragments from Zea mays. *Plant J.* **23**, 557–566 (2000).

99. Robertson, D. S. A possible dose-dependent inactivation of mutator (Mu) in maize. *Mol. Gen. Genet.* **191**, 86–90 (1983).
100. Walbot, V. & Rudenko, G. N. *MuDR/Mu* Transposable Elements of Maize. in *Mobile DNA II* (eds. Craig, N. L., Craigie, R., Gellert, M. & Lambowitz, A. M.) 533–564 (ASM Press, 2002).
101. Brown, W. E., Robertson, D. S. & Bennetzen, J. L. Molecular analysis of multiple mutator-derived alleles of the bronze locus of maize. *Genetics* **122**, 439–445 (1989).
102. Levy, A. A. *et al.* Developmental and genetic aspects of Mutator excision in maize. *Dev. Genet.* **10**, 520–531 (1989).
103. Levy, A. A. & Walbot, V. Molecular analysis of the loss of somatic instability in the bz2::mu1 allele of maize. *Mol. Gen. Genet.* **229**, 147–151 (1991).
104. Hershberger, R. J., Warren, C. A. & Walbot, V. Mutator activity in maize correlates with the presence and expression of the Mu transposable element Mu9. *Proc. Natl. Acad. Sci. U. S. A.* **88**, 10198–10202 (1991).
105. Bennetzen, J. L. The many hues of plant heterochromatin. *Genome Biol.* **1**, REVIEWS107 (2000).
106. Brown, J. & Sundaresan, V. Genetic study of the loss and restoration of Mutator transposon activity in maize: evidence against dominant-negative regulator associated with loss of activity. *Genetics* **130**, 889–898 (1992).
107. Stinard, P. S., Robertson, D. S. & Schnable, P. S. Genetic Isolation, Cloning, and Analysis of a Mutator-Induced, Dominant Antimorph of the Maize amylose extender1 Locus. *Plant Cell* **5**, 1555–1566 (1993).
108. Greene, B., Walko, R. & Hake, S. Mutator insertions in an intron of the maize knotted1 gene result in dominant suppressible mutations. *Genetics* **138**, 1275–1285 (1994).
109. Chen, C. H., Oishi, K. K., Kloeckener-Gruissem, B. & Freeling, M. Organ-specific expression of maize Adh1 is altered after a Mu transposon insertion. *Genetics* **116**, 469–477 (1987).
110. Bortiri, E. *et al.* ramosa2 encodes a LATERAL ORGAN BOUNDARY domain protein that determines the fate of stem cells in branch meristems of maize. *Plant Cell* **18**, 574–585 (2006).
111. Hirsch, C. D., Springer, N. M. & Hirsch, C. N. Genomic limitations to RNA sequencing expression profiling. *Plant J.* **84**, 491–503 (2015).
112. Chatterjee, M. & Martin, C. Tam3 produces a suppressible allele of the DAG locus of Antirrhinum majus similar to Mu-suppressible alleles of maize. *Plant J.* **11**, 759–771 (1997).
113. Raizada, M. N. & Walbot, V. The late developmental pattern of Mu transposon excision is conferred by a cauliflower mosaic virus 35S -driven MURA cDNA in transgenic maize. *Plant Cell* **12**, 5–21 (2000).
114. Hu, G., Yalpani, N., Briggs, S. P. & Johal, G. S. A porphyrin pathway impairment is responsible for the phenotype of a dominant disease lesion mimic mutant of maize. *Plant Cell* **10**, 1095–1105 (1998).
115. Liu, P., McCarty, D. R. & Koch, K. E. Transposon Mutagenesis and Analysis of Mutants in UniformMu Maize (Zea mays). in *Current Protocols in Plant Biology* (John Wiley & Sons, Inc., 2016).
116. Lawrence, C. J., Dong, Q., Polacco, M. L., Seigfried, T. E. & Brendel, V. MaizeGDB, the community database for maize genetics and genomics. *Nucleic Acids Res.* **32**, D393–7 (2004).

117. Zhou, P., Hirsch, C. N., Briggs, S. P. & Springer, N. M. Dynamic Patterns of Gene Expression Additivity and Regulatory Variation throughout Maize Development. *Mol. Plant* **12**, 410–425 (2019).
118. Monnahan, P. J. *et al.* Using multiple reference genomes to identify and resolve annotation inconsistencies. *BMC Genomics* **21**, 281 (2020).
119. Zhu, J., Kaeppler, S. M. & Lynch, J. P. Mapping of QTL controlling root hair length in maize (Zea mays L.) under phosphorus deficiency. *Plant Soil* **270**, 299–310 (2005).
120. Ewels, P. A. *et al.* The nf-core framework for community-curated bioinformatics pipelines. *Nat. Biotechnol.* **38**, 276–278 (2020).
121. Di Tommaso, P. *et al.* Nextflow enables reproducible computational workflows. *Nat. Biotechnol.* **35**, 316–319 (2017).
122. Krueger, F. *TrimGalore: A wrapper around Cutadapt and FastQC to consistently apply adapter and quality trimming to FastQ files, with extra functionality for RRBS data*. (Github).
123. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–360 (2015).
124. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
125. Oshlack, A., Robinson, M. D. & Young, M. D. From RNA-seq reads to differential expression results. *Genome Biol.* **11**, 220 (2010).
126. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
127. Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652 (2011).
128. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
129. Priyam, A. *et al.* Sequenceserver: A Modern Graphical User Interface for Custom BLAST Databases. *Mol. Biol. Evol.* **36**, 2922–2924 (2019).
130. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
131. Duvaud, S. *et al.* Expasy, the Swiss Bioinformatics Resource Portal, as designed by its users. *Nucleic Acids Res.* **49**, W216–W227 (2021).
132. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
133. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
134. R Core Team. *R: A Language and Environment for Statistical Computing*. https://www.r-project.org/ (2020).
135. Manoli, A., Sturaro, A., Trevisan, S., Quaggiotti, S. & Nonis, A. Evaluation of candidate reference genes for qPCR in maize. *J. Plant Physiol.* **169**, 807–815 (2012).
136. Brkljacic, J. & Grotewold, E. Combinatorial control of plant gene expression. *Biochim. Biophys. Acta* **1860**, 31–40 (2017).
137. Farré, G. *et al.* Engineering complex metabolic pathways in plants. *Annu. Rev. Plant Biol.* **65**, 187–223 (2014).
138. Yang, F., Ouma, W. Z., Li, W., Doseff, A. I. & Grotewold, E. Establishing the Architecture of Plant Gene Regulatory Networks. *Methods Enzymol.* **576**, 251–304 (2016).

139. Arda, H. E. & Walhout, A. J. M. Gene-centered regulatory networks. *Brief. Funct. Genomics* **9**, 4–12 (2010).
140. Brady, S. M. *et al.* A stele-enriched gene regulatory network in the Arabidopsis root. *Mol. Syst. Biol.* **7**, 459 (2011).
141. Taylor-Teeples, M. *et al.* An Arabidopsis gene regulatory network for secondary cell wall synthesis. *Nature* **517**, 571–575 (2015).
142. Ikeuchi, M. *et al.* A Gene Regulatory Network for Cellular Reprogramming in Plant Regeneration. *Plant Cell Physiol.* **59**, 765–777 (2018).
143. Yang, F. *et al.* A Maize Gene Regulatory Network for Phenolic Metabolism. *Mol. Plant* **10**, 498–515 (2017).
144. Walhout, A. J. M. What does biologically meaningful mean? A perspective on gene regulatory network validation. *Genome Biol.* **12**, 109 (2011).
145. Johnson, D. S., Mortazavi, A., Myers, R. M. & Wold, B. Genome-wide mapping of in vivo protein-DNA interactions. *Science* **316**, 1497–1502 (2007).
146. Kuo, M. H. & Allis, C. D. In vivo cross-linking and immunoprecipitation for studying dynamic Protein:DNA associations in a chromatin environment. *Methods* **19**, 425–433 (1999).
147. Bolduc, N. *et al.* Unraveling the KNOTTED1 regulatory network in maize meristems. *Genes Dev.* **26**, 1685–1690 (2012).
148. Eveland, A. L. *et al.* Regulatory modules controlling maize inflorescence architecture. *Genome Res.* **24**, 431–443 (2014).
149. Li, C. *et al.* Genome-wide characterization of cis-acting DNA targets reveals the transcriptional regulatory framework of opaque2 in maize. *Plant Cell* **27**, 532–545 (2015).
150. Morohashi, K. *et al.* A genome-wide regulatory framework identifies maize pericarp color1 controlled genes. *Plant Cell* **24**, 2745–2764 (2012).
151. Eisen, M. B., Spellman, P. T., Brown, P. O. & Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 14863–14868 (1998).
152. Huang, J., Vendramin, S., Shi, L. & McGinnis, K. M. Construction and Optimization of a Large Gene Coexpression Network in Maize Using RNA-Seq Data. *Plant Physiol.* **175**, 568–583 (2017).
153. Lu, X. *et al.* Gene-Indexed Mutations in Maize. *Mol. Plant* **11**, 496–504 (2018).
154. Gaut, B. S., Le Thierry d'Ennequin, M., Peek, A. S. & Sawkins, M. C. Maize as a model for the evolution of plant nuclear genomes. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 7008–7015 (2000).
155. Burdo, B. *et al.* The Maize TFome--development of a transcription factor open reading frame collection for functional genomics. *Plant J.* **80**, 356–366 (2014).
156. Zhou, P. *et al.* Meta Gene Regulatory Networks in Maize Highlight Functionally Relevant Regulatory Interactions. *Plant Cell* **32**, 1377–1396 (2020).
157. Schnable, P. S. *et al.* The B73 maize genome: complexity, diversity, and dynamics. *Science* **326**, 1112–1115 (2009).
158. Jiao, Y. *et al.* Improved maize reference genome with single-molecule technologies. *Nature* **546**, 524–527 (2017).
159. Liu, S., Dietrich, C. R. & Schnable, P. S. DLA-based strategies for cloning insertion mutants: cloning the gl4 locus of maize using Mu transposon tagged alleles. *Genetics* **183**, 1215–1225 (2009).
160. Wimalanathan, K., Friedberg, I., Andorf, C. M. & Lawrence-Dill, C. J. Maize GO Annotation-Methods, Evaluation, and Review (maize-GAMER). *Plant Direct* **2**, e00052 (2018).

161. Zhou, P. *et al.* Meta gene regulatory networks in maize highlight functionally relevant regulatory interactions. *Plant Cell* **32**, 1377–1396 (2020).
162. Andorf, C. M. *et al.* MaizeGDB update: new tools, data and interface for the maize model organism database. *Nucleic Acids Res.* **44**, D1195–201 (2016).
163. Franco-Zorrilla, J. M. *et al.* DNA-binding specificities of plant transcription factors and their potential to define target genes. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 2367–2372 (2014).
164. Perisic, O., Xiao, H. & Lis, J. T. Stable binding of Drosophila heat shock factor to head-to-head and tail-to-tail repeats of a conserved 5 bp recognition unit. *Cell* **59**, 797–806 (1989).
165. Grant, C. E., Bailey, T. L. & Noble, W. S. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017–1018 (2011).
166. Lund, J. Calculation of the representation factor and the associated probability. http://nemates.org/MA/progs/representation.stats.html.