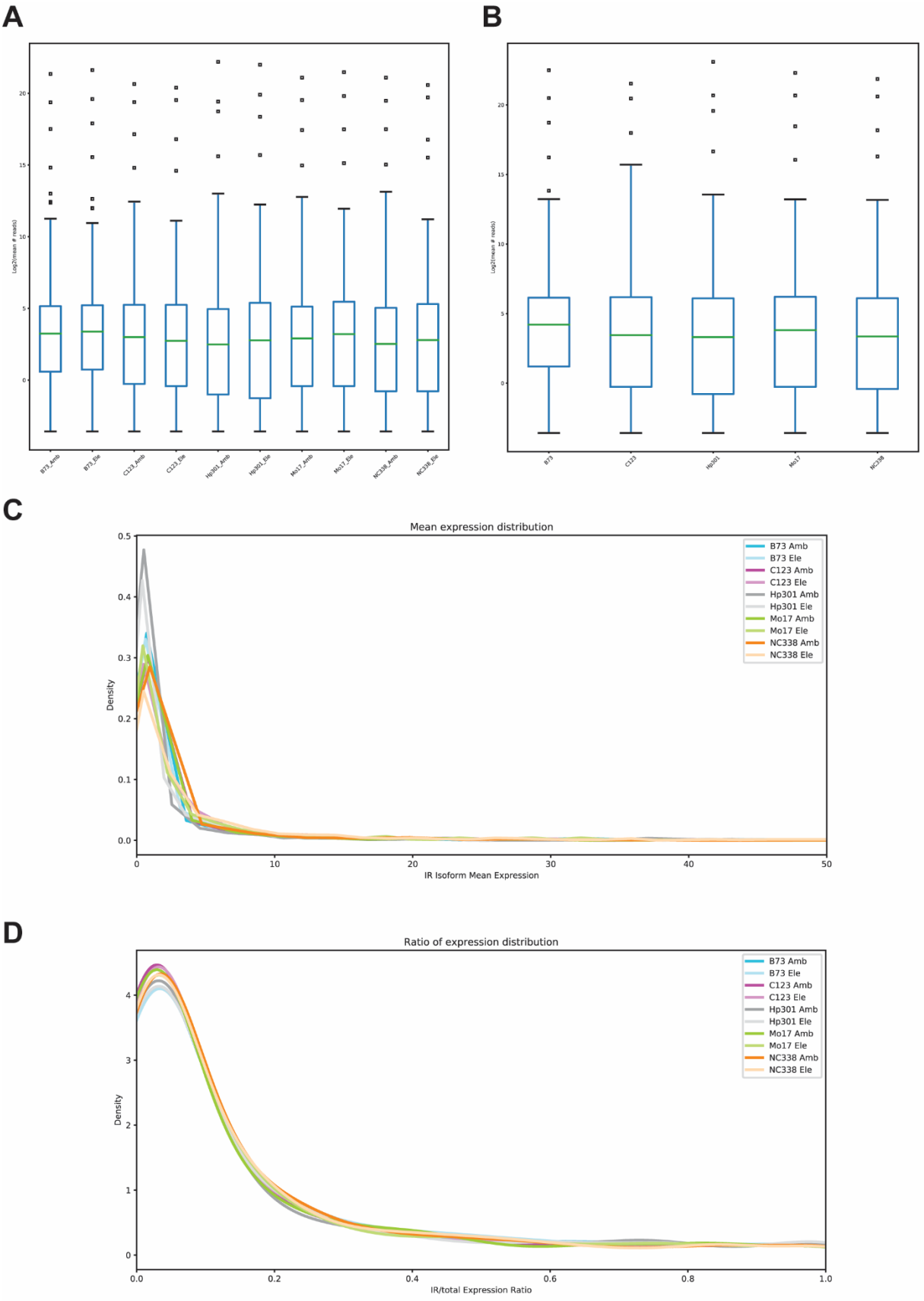
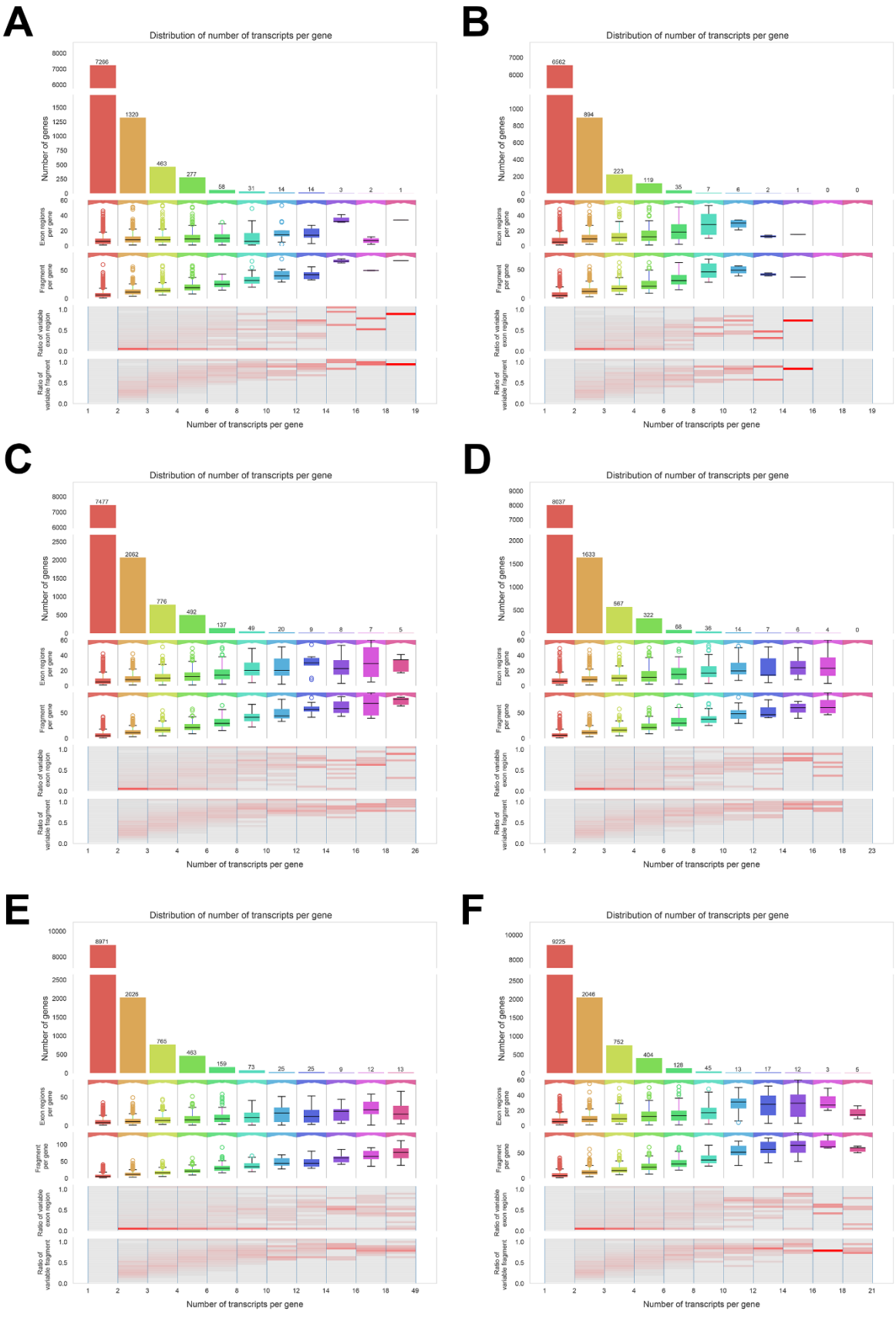
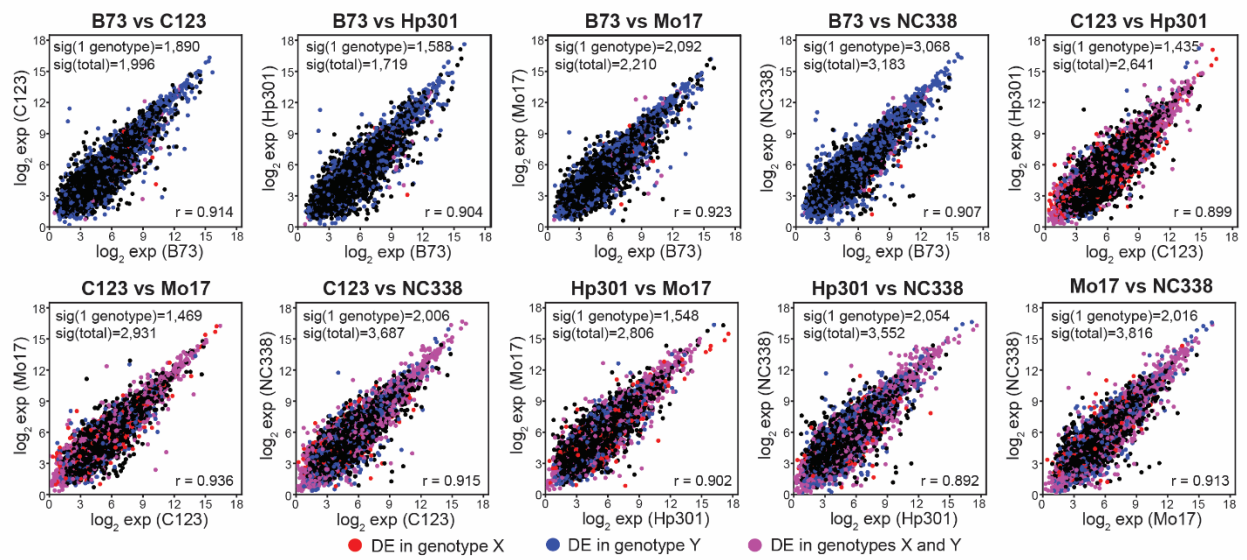

A
B
C
Mean expression distribution
Density
IR Isoform Mean Expression
B73 Amb
B73 Ele
C123 Amb
C123 Ele
Hp301 Amb
Hp301 Ele
Mo17 Amb
Mo17 Ele
NC338 Amb
NC338 Ele
D
Ratio of expression distribution
Density
IR/total Expression Ratio

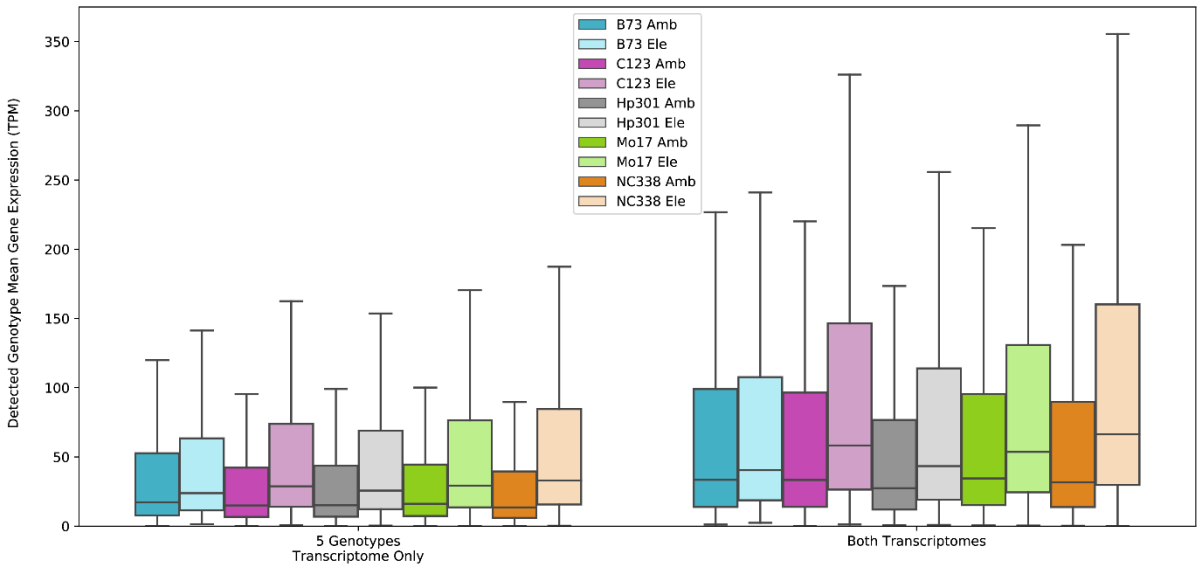
Supplementary Figure 1. The log2 scaled distribution of the mean number of reads associated with novel, fusion, or antisense loci for (A) each genotype – treatment combination and (B) each genotype summed across treatments. For annotated genes with 2 transcripts, (C) the distribution of mean expression (TPM) of transcripts with intron retention (IR) within each genotype and treatment sample, and (D) the distribution of the proportional contribution of the IR transcript expression to the total gene expression (ratio of IR expression to the total gene expression).


A
Distribution of number of transcripts per gene
Number of genes
Exon regions per gene
Fragment per gene
Ratio of variable exon region
Ratio of variable fragment
Number of transcripts per gene
B
Distribution of number of transcripts per gene
Number of genes
Exon regions per gene
Fragment per gene
Ratio of variable exon region
Ratio of variable fragment
Number of transcripts per gene
C
Distribution of number of transcripts per gene
Number of genes
Exon regions per gene
Fragment per gene
Ratio of variable exon region
Ratio of variable fragment
Number of transcripts per gene
D
Distribution of number of transcripts per gene
Number of genes
Exon regions per gene
Fragment per gene
Ratio of variable exon region
Ratio of variable fragment
Number of transcripts per gene
E
Distribution of number of transcripts per gene
Number of genes
Exon regions per gene
Fragment per gene
Ratio of variable exon region
Ratio of variable fragment
Number of transcripts per gene
F
Distribution of number of transcripts per gene
Number of genes
Exon regions per gene
Fragment per gene
Ratio of variable exon region
Ratio of variable fragment
Number of transcripts per gene

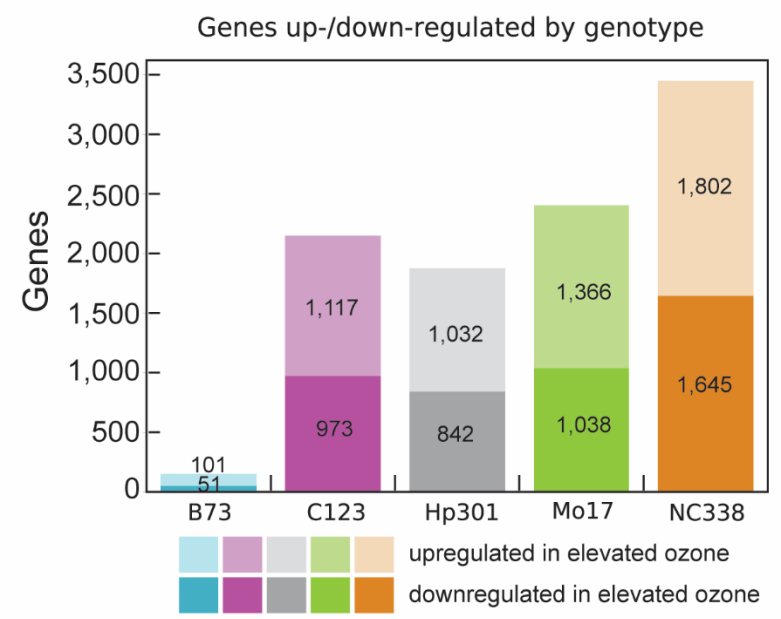
Supplementary Figure 2. Frequency of transcripts per gene and, for each bin, the number of exon regions/fragments per gene and the proportion of varying exon regions/fragments in the Wang et al. 2018 data. Transcripts with unique splice junctions for the individual maize tissue transcriptomes from (A) leaf, (B) silk, (C) pericarp, (D) bract, (E) shoot and (F) seedling. In the single tissue leaf transcriptome, the number of expressed transcripts (13,580 transcripts, 9,451 genes) is lower than what is annotated for the reference. Approximately 77% of genes have a single expressed transcript.



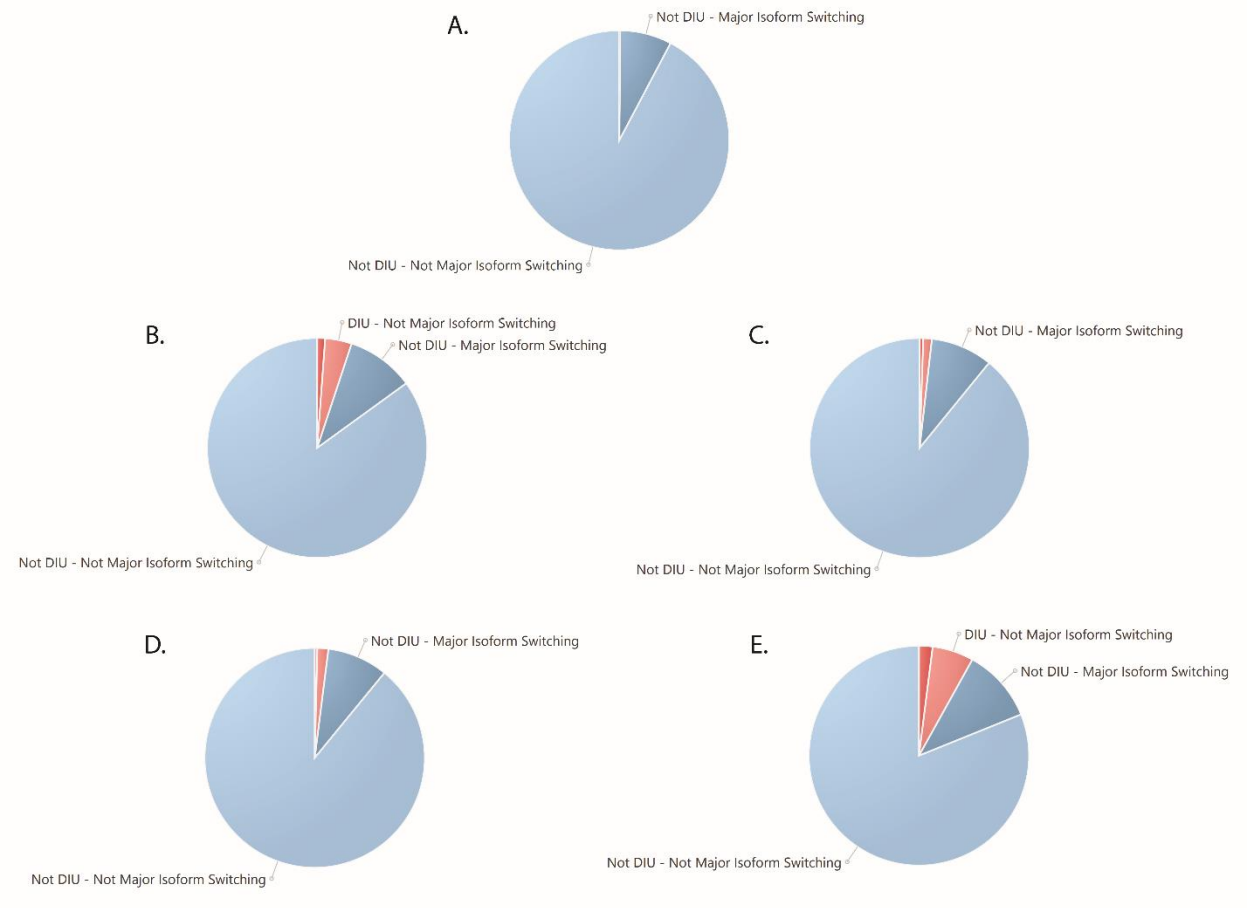
Supplementary Figure 3. Differentially expressed genes**.** A scatterplot for $\log_2$(TPM) for each pair of genotypes in the ambient condition. Genes in each plot are colored based on differential gene expression between ambient and ozone: (1) red genes are differentially expressed in the genotype on the x-axis; (2) genes colored blue are differentially expressed in the genotype on the y-axis; (3) genes differentially expressed in both genotypes are colored purple; (4) genes not differentially expressed in either genotype are colored black. The relatively small transcriptional response of B73 can be seen in the first 4 panels (B73 on x-axis): the proportion of genes colored blue compared to genes colored red or purple is higher for plots with B73 that in plots without B73. Other genotype combinations show a larger degree of a shared response (e.g., a greater number of genes colored purple). We note that there is an even distribution of differentially expressed genes across the range of expression in the ambient condition.



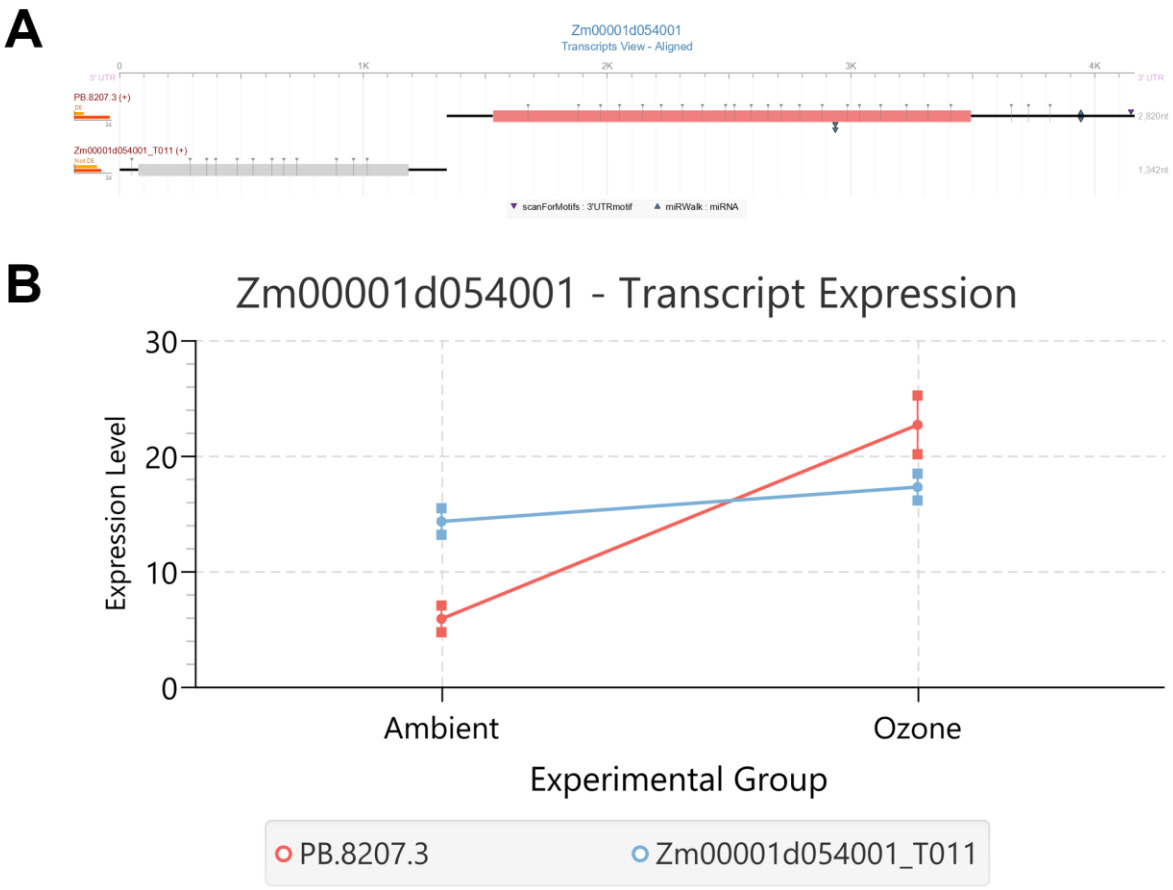
Supplementary Figure 4. For each genotype – treatment combination, mean gene expression (TPM) of genes detected only in the maize leaf transcriptome is shown on the left (n=5269) and genes detected in both the maize leaf transcriptome and in the B73 leaf transcriptome is shown on the right (n=7249).



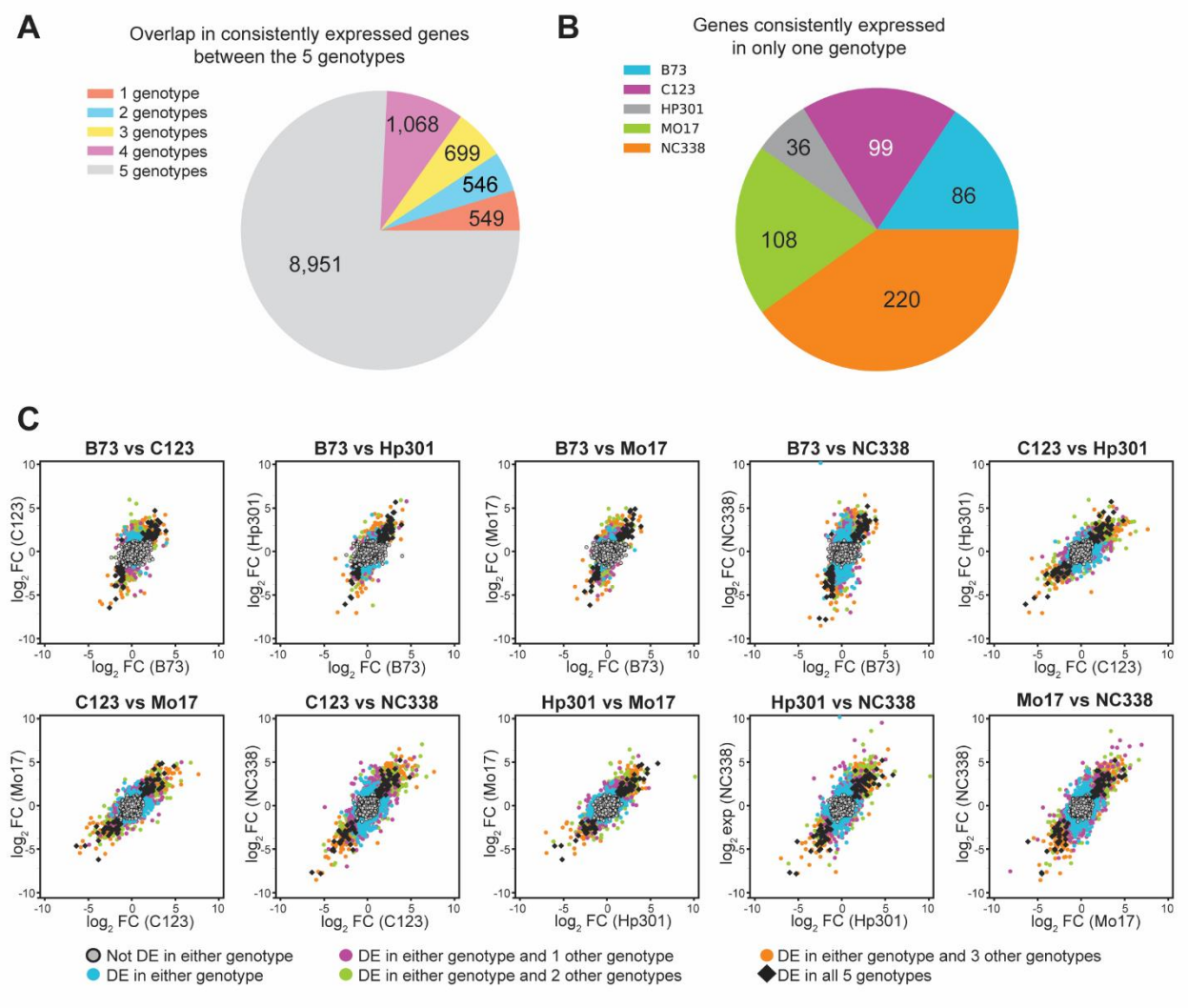
Supplementary Figure 5. For all genotypes, differential expression was roughly balanced between over- and under-expression with B73 showing a relatively minor response and NC338 showing the most extreme response to ozone stress.



Supplementary Figure 6. Summary of tappAS [1] differential isoform usage (DIU) results for the 5 genotypes in this study. The number of gene with DIU and major isoform switching (n, dark red) for (A) B73 (n=1, 2874 genes tested), (B) C123 (n=34, 2941 genes tested), (C) Hp301 (n=14, 2696 genes tested), (D) Mo17 (n=11, 2961 genes tested), and (E) NC338 (n=59, 2966 genes tested). Few of the genes with DIU and major isoform switching are shared across multiple genotypes. Only 15 genes are shared by 2 genotypes and 1 gene (Zm00001d054001) is shared by Hp301, Mo17, and NC338.



Supplementary Figure 7. An example gene (Zm00001d054001) identified with DIU and major isoform switching by tappAS [1] in Hp301, Mo17 and NC338 genotypes. Transcript models are shown in (A) and transcript expression in ambient and ozone conditions for genotype NC338 in (B).



Supplementary Figure 8. Maize transcriptome expression. (A) Overlap of genes consistently expressed across genotypes. (B) Number of genes consistently expressed in only a single genotype. (C) A scatterplot of the $\log_2$(Fold Change) for each indicated pair of genotypes. Genes in each scatterplot are colored based on the number of genotypes for which that gene is differentially expressed between ambient and ozone: (1) genes colored gray with black outline are not differentially expressed in the genotypes indicated on the axes; (2) genes colored pink are differentially expressed in either genotype indicated on the axes; (3) genes differentially expressed in either genotype indicated on the axes plus 1 additional genotype are colored blue; (4) genes differentially expressed in either genotype indicated on the axes plus 2 additional genotypes are colored yellow; (5) genes differentially expressed in either genotype indicated on the axes plus 3 additional genotypes are colored purple; (6) genes differentially expressed in all 5 genotypes are shown as gray diamonds.

Supplementary Table 1. PacBio transcriptome mapping and performance metrics. High-quality transcripts from each sequenced PacBio library (genotype × ozone treatment) were mapped to 3 reference genomes: (i) B73 RefGen_v4 [2], (ii) Mo17 Yan [3] and (iii) Mo17 Cau [4]. For each row parameter x sample combination, PacBio read numbers generated via mapping (minimap2, v2.12, [5]) the PacBio polished transcripts to the B73, Mo17 Yan and Mo17 Cau references, respectively, are separated by forward slashes. Cupcake ToFU2 (version 20180629) supporting scripts were used to generate unique isoforms. The number if PacBio transcripts 'ignored' by ToFU2 due to low coverage, low identify or non-mapping are indicated for each reference. Results from the QC function of SQANTI [6] are noted for each library and for the combined chained samples: FSM, full-splice match; ISM, incomplete-splice match; NIC, novel-in-catalog; NNC, novel-not-in-catalog. Chained transcriptome SQANTI filtering and curation are also included.

Supplementary Table 2. Quality control information for PacBio read processing.

Supplementary Table 3. Novel and antisense loci with at least one read in any replicate in B73 only, including the 3 novel loci with an average of at least 5 reads across replicates in both treatments (novelGene_372, novelGene_344, and novelGene_769). Read counts for each genotype by treatment combination is given.

Supplementary Table 4. Genes consistently expressed only in (A) ozone or (B) ambient. The flag_*genotype_condition* variables are 0/1 binary indicator flags for detection: the flag is equal to 1 if TPM > 5 in > 50% of the replicates for the given genotype - condition. Flag_all_on_*condition* is a 0/1 binary indicator variable for whether a given gene is expressed (the flag is equal to 1) in all genotypes for the given condition. Flag_detect_*genotype* = 1 if expressed in 1 of the 2 conditions. The tables also include the mean expressions for the indicated genotype – condition.

Supplementary Table 5. Genes with evidence of differential isoform usage and major isoform switching in at least one genotype. The DIU_qval_*genotype* columns contain the q-value for the indicated DIU test in tappAS. The flag_DIU_*genotype* variables = 1 if the tappAS DIU test was significant (q-value < 0.05) for the indicated genotype. The flag_DIU_majorisoformSwitch_*genotype* variables = 1 for the indicated genotype if the DIU test was significant and the test showed a major isoform switch. The total_usage_change_*genotype* columns contain the total change in expression across transcripts for the indicated genotype. The mean_TPM_*condition_genotype* columns contain the mean TPM values for the indicated genotype – condition. The flag_detect_DIU_majorIsoformSwitch_*genotype* columns = 1 if the gene is significant in DIU test, contains a major switch and is detected in the given genotype.

Sum_detect_DIU_majorisoformswitch is the sum of the flag_detect_DIU_majorIsoformSwitch_*genotype* flags.

Supplementary Table 6. GO enrichment results. The flag_detect_DE_all5 variable = 1 for genes differentially expressed in all 5 genotypes. The flag_detect_DE_*genotype* variables = 1 for genes differentially expressed in the indicated genotype. The flag_detect_DE_all_noB73 variable = 1 for genes differentially expressed in C123, Hp301, Mo17, and NC338, but not B73. The flag_B73Zero_restUp variable = 1 for genes with a fold change between ambient and ozone of 0 for B73 and > 0 for all other genotypes in genes differentially expressed in at least one genotype.

Supplementary File 1. Functional annotation (in GFF-like format) mapped to the B73 genome of the curated maize leaf transcriptome generated from the data described in this study.

Supplementary File 2. GTF of the curated maize leaf transcriptome mapped to the B73 genome.

Supplementary File 3. The complete expression matrix of the curated maize leaf transcriptome. Expression values (TPM) for each transcript are given in the indicated *genotype_plant_chamber_condition* columns. The flag_*genotype_condition* variables are 0/1 binary indicator flags for detection, where the flag is equal to 1 if TPM > 5 in > 50% of the replicates. The mean TPM for each genotype–condition is given in the mean_*gentoype_condition* columns.

Supplementary File 4. Gene level differential expression results for the maize leaf transcriptome. The flag_*genotype_condition* variables are 0/1 binary indicator flags for detection where the flag is equal to 1 if at least 1 transcript for the indicated gene was detected. The sum_flag variable is the sum of the flag_*genotype_condition* variables for a given gene. The DE_pval_*genotype* columns contain the p-value for the test of DE. The flag_DE_*genotype* variable = 1 if the test of DE was significant (p-value < 0.05). The Log2FC_*genotype* columns contain the TPM fold change between ambient and ozone conditions. The mean TPM for each genotype–condition is given in the mean_TPM_*genotype_condition* columns. The flag_detect_DE_*genotype* variables = 1 for genes differentially expressed and detected in the indicated genotype. The flag_detect_DE_all5 variable = 1 for genes differentially expressed and detected in all 5 genotypes. GO IDs and GO terms are included for each gene.

Supplemental Methods

Long-read library preparation and sequencing: Total RNA was isolated followed by full-length cDNA synthesis using the SMARTer PCR cDNA Synthesis kit. Two cDNA reactions were performed for each sample. Single molecule SMRTbell libraries were constructed from each double-stranded cDNA sample using the IsoSeq protocol followed by PCR optimization and ELF size selection. Libraries were sequenced and processed on the PacBio SEQUEL platform using 1 LR SMRT cell per sample and SMRT Link 6.0.0 chemistry to generate raw subreads for each sample.

Long-read pre-processing: The IsoSeq3 pipeline (v3.0.0, https://github.com/PacificBiosciences/IsoSeq3) was conducted on each individual library starting with subread .bam files. CCS (v3.1.0) was used to process subreads into single consensus reads. Consensus reads for each sample have, on average, approximately 3 sub-reads, indicating successful multiple passes during sequencing and enabling error correction. Primers were removed using lima (v1.7.1). IsoSeq3 cluster was used to trim polyA tails, remove false concatemers and cluster full-length reads by sequence similarity. Consensus ('polished') sequences for each cluster were generated using IsoSeq3 polish. The expected error rate after this step is less than < 2% [7]. To this point no references were used. The quality of the long-read data was high as determined by high numbers of sub-reads and the relatively few reads removed downstream read counts during pre-processing (Supplementary Table 1). Clustering and polishing of trimmed reads yielded 24,000 to 44,000 high quality individual long-reads per library (Supplementary Table 2).

Long-read mapping polished clusters: Polished transcript sequences were aligned to reference genomes B73 (Ensembl version 41, ftp://ftp.ensemblgenomes.org/pub/plants/release-41/fasta/zea_mays/, [2]), Mo17 Yan (https://ftp.maizegdb.org/MaizeGDB/FTP/Mo17-YAN/ , [3]) and Mo17 Cau (https://ftp.maizegdb.org/MaizeGDB/FTP/Mo17-CAU/, [4]) using minimap2 (v2.12, [5]) and parameters recommended for PacBio IsoSeq3 processing (https://github.com/Magdoll/cDNA_Cupcake/wiki/Best-practice-for-aligning-Iso-Seq-to-reference-genome:-minimap2,-deSALT,-GMAP,-STAR,-BLAT). We refer to these as the B73, Mo17 Yan and Mo17 Cau references. Map positions were used to identify identical isoforms within each library. The IsoSeq3 supporting script in Cupcake ToFU2 (version 20180629, https://github.com/Magdoll/cDNA_Cupcake, collapse_isoforms_by_sam.py) was used to obtain a set of unique, full-length and high-quality transcripts for each library. Additional IsoSeq3 supporting scripts in Cupcake ToFU2 (summarize_sample_GFF_junctions.py and scrub_sample_GFF_junctions.py) were used to summarize junctions across all samples and filter junctions for probable artifacts based on presence in the reference or long-read support across all samples.

Long-read chaining multiple samples: The high quality PacBio transcripts from all samples were combined, or chained, into a unique set of isoforms using IsoSeq3 supporting script, chain_samples.py, from Cupcake ToFU2. Isoforms present in at least one library were identified

and assigned a common identifier if contained within more than one library. This protocol was carried out on PacBio transcripts mapped to the B73 and Mo17 Yan references separately, yielding 2 chained sets of isoforms (Supplementary Table 1). SQANTI quality control was run on each chained set to return an annotated transcriptome. There were fewer unmapped transcripts, a higher number of annotated transcripts and genes, and a lower number of novel genes when the chained samples were mapped to the B73 reference than to the Mo17 Yan reference (Fig. 1D, Supplementary Table 1). Of the annotated transcripts, there were more full-splice matches mapped to the B73 reference for all but the Mo17 samples (where the difference was minor). This is evidence that the B73 genome has a higher quality annotation than the Mo17 Yan genome. Chaining identifies transcripts present in multiple libraries and assigns an additional arbitrary PacBio identifier ('super' PacBio identifiers). Super PacBio identifiers should represent a transcript. However, small differences in sequence will result in different Super PacBio identifiers. In addition, a few sequenced transcripts are expected to be technical artifacts [6].

Sequenced transcript quality evaluation: We used SQANTI (sqanti_qc.py, [6]) to evaluate the transcripts generated from individual libraries (pre-chaining) and all libraries combined (post-chaining) mapped to both references (B73 or Mo17 Yan). SQANTI characterizes transcripts based on correspondence to reference transcripts. A full-splice match (FSM) is a transcript where all splice junctions match exactly to an annotated reference transcript. There may be some variation in 3' and especially 5' regions due to improper annotation, degradation, or a technical artifact of library preparation [6]. Incomplete-splice matches (ISM), on the other hand, have junctions in the same order as a reference transcript, but may be missing some on the 5' and/or 3' end of the transcript. Some transcripts can contain novel combinations of annotated junctions, called novel in catalog (NIC), while others may have entirely new junctions within an annotated gene, called novel not in catalog (NNC). Characterizations of transcripts within novel genes include genic genomic, genic intron, intergenic, antisense, and fusion based on where they are aligned compared to the annotated reference genes [6] (Supplementary Table 1).

Evaluation of reference genomes: References for the B73 and Mo17 form a basis for the quality control examination of long-read transcripts for both of these genotypes. By comparing long-read transcripts to their corresponding reference genomes (the B73 reference [2]; Mo17 Yan reference [3] or the Mo17 Cau reference [4]), we gain insight into potential technical artifacts. Comparing long-read transcripts from B73 to the Mo17 reference and *vice versa* allows us to estimate the divergence of the transcriptomes. We mapped all sequenced reads from each library to all three possible genomic references (Fig. 1A, Supplementary Table 1). Long-read transcripts that map to their corresponding reference genome with low coverage or with low identity were also identified. There are only a few unmapped long-read transcripts (Fig. 1B, Supplementary Table 1). For B73 samples mapped to the B73 reference, there were 12 unmapped sequenced transcripts: 5 in the ambient condition and 7 in ozone. Only 2 of these 12 transcripts were greater than 1kb in length. These transcripts were further examined by comparing to the B73 reference genome with BLAST [8]. Three of the 12 hit to the same small (61bp) region of the B73 genome while the remaining 9 sequences did not return a hit. These results indicate these 12 transcripts from B73 are likely to

reflect technical artifacts. Similarly, for the 3 Mo17 samples (note that the Mo17 ozone library was sequenced twice, once for optimization of sequencing parameters and then again with all sample libraries) there were 363 sequenced transcripts unmapped to the Mo17 Yan reference (79 in the ambient ozone library and 284 in the 2 elevated ozone libraries) and 50 unmapped to the Mo17 Cau reference (14 in the ambient ozone library and 36 in the 2 elevated ozone libraries) (Fig. 1B, Supplementary Table 1). The difference in mappability between the Mo17 Yan and Mo17 Cau reference genomes indicates that the Mo17 Cau genome is potentially a better genome assembly. We evaluated the 50 Mo17 sequenced transcripts that did not map to the Mo17 Cau reference comparing them to the Mo17 Cau reference using BLAST. Similar to the results with B73, the alignments returned were partial hits, many containing mismatches and gaps. Taken together, this indicates that less than 1% of the polished clusters for the 4 samples were likely technical artifacts.

The B73 genome as reference: We used the program SQANTI [6] to examine the quality of the sequenced transcriptomes. SQANTI output includes classifications that describe the exon structure of the sequenced transcripts compared to the annotated transcriptome. Full-splice match (FSM) and incomplete-splice match (ISM) classifications reflect whether the transcript is present, in full or in part, in the reference transcriptome. Novel in catalog (NIC) transcripts contain known splice junctions or new splice junctions from known donors and acceptors, while the novel not in catalog (NNC) transcripts contain novel donors and/or acceptors. We applied SQANTI quality control to the transcripts sequenced from each individual PacBio library mapped to all three reference genomes. For all but Mo17, the highest FSM numbers were identified using the B73 reference (Figure 1C). The relatively few annotated transcripts in the Mo17 Cau reference led us to focus on the B73 annotated genome (Fig 1C). There were several hundred previously un-annotated transcripts in genic regions in each library, indicating the presence of potentially novel gene sequences in the transcriptome of the leaf (Supplementary Table 1).

Sequences that do not map to B73 represent technical artefacts: Sequenced transcripts generated from Mo17, C123, NC338 and Hp301 were mapped to the B73 genome. Transcripts that did not map (Figure 1B) could represent genes missing from the B73 reference genome, genes that have significantly diverged between the genotypes, or gaps in the assembly. We compared mapping between B73 and Mo17 where the reference genomes were compared in detail [9]. For B73, there were ~350 transcripts (from either ambient or elevated ozone) that did not map to the Mo17 Yan reference (Figure 1B). Only 63 B73 transcripts did not map to the Mo17 Cau reference genome, consistent with the hypothesis that the two Mo17 genome assemblies differ in quality. A similar number of Mo17 transcripts did not map to the Mo17 Cau reference (n=50). Additional support for genome quality contributing to the mappability of the sequenced transcripts is the observation that sequenced transcripts from Mo17 mapped better to the B73 reference genome than to either Mo17 reference genome, as only 33 Mo17 transcripts were unmapped to the B73 reference (Fig. 1B, Supplementary Table 1) and a very similar, low number of sequenced transcripts were found unmapped to the B73 reference genome for the other three genotypes (C123, Hp301 and NC338). The NC338 ozone library had the most, but still a very small percentage, of transcripts unmapped to the B73 reference at 120 (Fig. 1B, Supplementary Table 1). As an additional check, we combined all libraries after the initial read processing but prior to Iso-seq3 clustering and mapping to the B73 reference. After mapping the combined polished transcripts to the B73 reference, we

collapsed all reads together. All but 250 of the resulting collapsed transcripts mapped to B73. BLAST results for these unmapped transcripts revealed partial hits, many with mismatches and gaps, suggesting these 250 transcripts were technical artifacts. We did not find evidence supporting the presence of genes expressed in the leaf that diverge in presence/absence variation between the genotypes.

SQANTI filtering: RNA-seq short reads were mapped to the PacBio transcripts with the program STAR (2.7.0b, [10]) and junction coverage estimated from the resulting alignments. The input expression matrix was obtained with RSEM running the Bowtie aligner (version 1.2.28, [11, 12]). Data for subsequent filtering was prepared using the sqanti_qc.py script with junction and transcript expression matrices included as input. The SQANTI filtering function (sqanti_filter.py, [6]) was used to identify and remove low-quality transcripts and sequencing artifacts from the chained samples mapped to B73. Transcripts with insufficient short-read support, non-canonical junctions, or evidence of a possible reverse transcriptase switching event were noted and removed from further analysis [6]. After SQANTI filtering there were 19,251 reference transcripts detected in the long-read data (FSM or ISM). An additional 14,016 transcripts contained within a known gene but not previously annotated as a transcript (NIC or NNC) for a total of 33,267 transcripts from 12,702 known genes expressed in the leaf. We identified 1,531 additional genes that were not annotated in the B73 reference.

Curation of filtered transcripts: The chained library samples were filtered using SQANTI and curated to remove redundancies (where multiple sequenced transcripts map to the same reference transcript identifier due to variation in the 3'UTR or 5'UTR). For the PacBio transcripts that matched to a known B73 reference transcript (FSM or ISM), the longest B73 reference transcript was selected. For novel transcripts with partial matches in annotated genes (NIC or NNC), we removed unspliced fragments from multi-exon reference genes. Reads associated with putative novel and fusion loci and those classified as "genic" (i.e., that lack any splice junctions/donors/acceptors that match those present in the reference annotation) were also removed as these are unlikely to represent complete transcripts. Transcripts were evaluated for redundant splice junctions and the longest transcript for each unique set of splice junctions was selected. The GTF for the curated maize leaf transcriptome can be found in Supplementary File 2. In brief, there were 31,388 transcripts with unique splice junctions in 12,604 annotated genes expressed in the leaf (Supplementary File 2).

Quantification of expression: Curated PacBio gene and transcript expression values were estimated by short-read quantification using RSEM running the Bowtie aligner [11, 12]. Expected counts and TPM (transcripts per kilobase million) values were extracted from the alignments for PacBio genes and PacBio transcripts. Single-end (R1-R2 merged reads) and paired-end counts were summed to give final values.

Database	Link
UniProt (Swiss-Prot)	https://www.uniprot.org/uniprot/?query=reviewed:yes%20AND%20organism:%22Zea%20mays%20[4577]%22#
UniProt (TrEMBL)	https://www.uniprot.org/uniprot/?query=reviewed:no%20AND%20organism:%22Zea%20mays%20[4577]%22#
RefSeq	ftp://ftp.ncbi.nlm.nih.gov/genomes/Zea_mays/protein/protein.fa.gz
Ensembl	ftp://ftp.ensemblgenomes.org/pub/plants/release-41/fasta/zea_mays/pep/Zea_mays.B73_RefGen_v4.pep.all.fa.gz

1. de la Fuente L, Arzalluz-Luque Á, Tardáguila M, del Risco H, Martí C, Tarazona S, Salguero P, Scott R, Alastrue-Agudo A, Bonilla P, et al: **tappAS: a comprehensive computational framework for the analysis of the functional impact of differential splicing.** *bioRxiv* 2019**:**690743.
2. Jiao Y, Peluso P, Shi J, Liang T, Stitzer M, Wang B, Campbell M, Stein J, Wei X, Chin C, et al: **Improved maize reference genome with single-molecule technologies.** *Nature* 2017, **546:**524-+.
3. Yang N, Xu X, Wang R, Peng W, Cai L, Song J, Li W, Luo X, Niu L, Wang Y, et al: **Contributions of Zea mays subspecies mexicana haplotypes to modern maize.** *Nature Communications* 2017, **8**.
4. Sun S, Zhou Y, Chen J, Shi J, Zhao H, Zhao H, Song W, Zhang M, Cui Y, Dong X, et al: **Extensive intraspecific gene order and gene structural variations between Mo17 and other maize genomes.** *Nature Genetics* 2018, **50:**1289-+.
5. Li H: **Minimap2: pairwise alignment for nucleotide sequences.** *Bioinformatics* 2018, **34:**3094-3100.
6. Tardaguila M, de la Fuente L, Marti C, Pereira C, Pardo-Palacios FJ, Del Risco H, Ferrell M, Mellado M, Macchietto M, Verheggen K, et al: **SQANTI: extensive characterization of long-read transcript sequences for quality control in full-length transcriptome identification and quantification.** *Genome Res* 2018.
7. Weirather JL, de Cesare M, Wang Y, Piazza P, Sebastiano V, Wang X-J, Buck D, Au KF: **Comprehensive comparison of Pacific Biosciences and Oxford Nanopore Technologies and their applications to transcriptome analysis.** *F1000Research* 2017, **6:**100-100.
8. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *Journal of Molecular Biology* 1990, **215:**403-410.
9. Hollick JB, Dorweiler JE, Chandler VL: **Paramutation and related allelic interactions.** *Trends Genet* 1997, **13:**302-308.
10. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR: **STAR: ultrafast universal RNA-seq aligner.** *Bioinformatics* 2013, **29:**15-21.

11. Li B, Dewey CN: **RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome.** *BMC Bioinformatics* 2011, **12:**323.
12. Langmead B, Trapnell C, Pop M, Salzberg S: **Ultrafast and memory-efficient alignment of short DNA sequences to the human genome.** *Genome Biology* 2009, **10**.