


RESEARCH

Open Access



# Analysis of transcriptomic differences between NK603 maize and near-isogenic varieties using RNA sequencing and RT-qPCR

Sina-Elisabeth Ben Ali<sup>1,2</sup> , Agnes Draxler<sup>1,2</sup>, Diana Poelzl<sup>1</sup>, Sarah Agapito-Tenfen<sup>3</sup>, Rupert Hochegger<sup>1</sup>, Alexander G. Haslberger<sup>2</sup> and Christian Brandes<sup>1\*</sup>

## Abstract

**Background:** The insertion of a transgene into a plant organism can, in addition to the intended effects, lead to unintended effects in the plants. To uncover such effects, we compared maize grains of two genetically modified varieties containing NK603 (AG8025RR2, AG9045RR2) to their non-transgenic counterparts (AG8025conv, AG9045conv) using high-throughput RNA sequencing. Moreover, in-depth analysis of these data was performed to reveal the biological meaning of detected differences.

**Results:** Uniquely mapped reads corresponded to 29,146 and 33,420 counts in the AG8025 and AG9045 varieties, respectively. An analysis using the R-Bioconductor package EdgeR revealed 3534 and 694 DEGs (significant differentially expressed genes) between the varieties AG8025RR2 and AG9045RR2, respectively, and their non-transgenic counterparts. Furthermore, a Deseq2 package revealed 2477 and 440 DEGs between AG8025RR2 and AG9045RR2, respectively, and their counterparts. We were able to confirm the RNA-seq results by the analysis of two randomly selected genes using RT-qPCR (reverse transcription quantitative PCR). PCA and heatmap analysis confirmed a robust data set that differentiates the genotypes even by transgenic event. A detailed analysis of the DEGs was performed by the functional annotation of GO (Gene Ontology), annotation/enrichment analysis of KEGG (Kyoto Encyclopedia of Genes and Genomes) ontologies and functional classification of resulting key genes using the DAVID Bioinformatics Package. Several biological processes and metabolic pathways were found to be significantly different in both variety pairs.

**Conclusion:** Overall, our data clearly demonstrate substantial differences between the analyzed transgenic varieties and their non-transgenic counterparts. These differences indicate that several unintended effects have occurred as a result of NK603 integration. Heatmap data imply that most of the transgenic insert effects are variety-dependent. However, identified key genes involved in affected pathways of both variety pairs show that transgenic independent effects cannot be excluded. Further research of different NK603 varieties is necessary to clarify the role of internal and external influences on gene expression. Nevertheless, our study suggests that RNA-seq analysis can be utilized as a tool to characterize unintended genetic effects in transgenic plants and may also be useful in the safety assessment and authorization of genetically modified (GM) plants.

**Keywords:** GMO, NK603, Transcriptomics, Roundup, RNA-seq

## Background

Herbicide-tolerant crops such as RR (Roundup Ready)-maize and RR-soybeans were first introduced approximately 25 years ago and now comprise the majority of cultivated GM crops worldwide [43]. In this study,

\*Correspondence: christian.brandes@aon.at

<sup>1</sup> Institute for Food Safety, Austrian Agency for Health and Food Safety, Spargelfeldstrasse 191, 1220 Vienna, Austria

Full list of author information is available at the end of the article

Roundup Ready maize (NK603) was analyzed. The NK603 transgene consists of two cassettes, which were inserted by microparticle bombardment. Both cassettes contain a 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS) gene obtained from the soil bacterium *Agrobacterium* sp. strain CP4. In the first cassette, the gene is regulated by the rice (*Oryza sativa*) actin promoter. The second cassette has nearly the same composition as the first one, but it is regulated by an enhanced 35S promoter obtained from the cauliflower mosaic virus. The CP4-EPSPS transgene confers plant tolerance to the herbicide glyphosate, commercially sold as Roundup®. The function of the EPSPS protein, its toxicity and allergenicity as well as metabolites have been examined in several projects [9, 29, 39, 70]. In the field of genetic engineering, there is still a lack of clarity about transgene integration mechanisms and about which biological processes take place in the plant after successful transgene integration [46, 74]. Since DNA integration can lead to gene disruptions, DNA rearrangements, or the production of new proteins [71], the insertion of a transgene into a plant organism can, in addition to the intended effects (e.g., herbicide resistance), lead to unintended effects in the plants [38, 50, 74]. Such unintended effects need to be identified and evaluated to detect and prevent possible adverse effects [28].

For the authorization and approval of GM crops, it is crucial to avoid the occurrence of unintended effects. Thus, investigations of GM crops on a molecular biology level are essential to enable a deeper understanding of plant gene interactions. Currently, unintended effects are primarily investigated by targeted approaches, in which GM crops are compared to their near-isogenic non-GM counterparts. Molecular, compositional, phenotypic, and agronomic analyses are performed in order to identify similarities and differences between the crops. Significant differences between a GM crop and its conventional counterpart indicate an unintended effect, which requires further investigation [30]. Targeted approaches have the disadvantage that important differences can be overlooked, as only selected features or chemical/nutritional compounds are investigated. Untargeted omics approaches, e.g., the analysis of gene expression in its entirety, enable the identification of broad coordinated trends that cannot be discerned by targeted approaches [54]. Thus, new non-targeted profiling approaches are an appropriate tool to complement current investigations regarding unintended effects of transgenic plants [31, 63, 71].

There are several studies investigating unintended effects in GM crops using different omics approaches and microarrays. In these studies, it was found that transgenic inserts of different species of plants might affect the

overall expression of other endogenous genes [1, 2, 7, 8, 12, 17, 19, 32, 35, 40, 45, 49, 60, 84, 90]. However, several authors of these studies mention that the changes in gene expression and protein distribution caused by genetic modification were smaller than those caused by environmental factors or natural variations. In addition, no differences were found in some of these studies. For instance, [21] and 2009, compared grown MON810 maize varieties with some comparable varieties using microarrays under in vitro conditions and conventional agricultural field conditions. In these studies, no gene was found to be significantly differentially expressed in any of the variety pairs tested under in vitro or conventional field conditions [20, 21].

Differences in lignin content between MON810 maize and comparators were reported by Saxena and Stotzky as well as Poerschmann [65, 73]. Herrero et al. detected significant differences in enantiomeric amino acid composition between two MON810 varieties and comparators but not in a third tested variety pair (MON810 vs. non-GM) [41]. Agapito et al. conducted a comparative analysis of MON810 maize and comparators grown under different agroecosystem conditions in Brazil using two-dimensional gel electrophoresis combined with mass spectrometry, which identified 32 differentially expressed proteins [1]. La Paz et al. performed RNA sequencing on MON810 maize and comparators, which revealed 140 differentially expressed genes. A slight, but significant, delay in seed and plant maturation of MON810 maize plants was also observed and, thus, used as an explanation for the detected differences [49]. [36].

Unintended effects in NK603 maize were described by several authors [2, 8, 12, 37, 40, 58]. Agapito-Tenfen et al. analyzed two maize events, NK603 and MON89034, as single events and as stacked varieties. Proteomics was used for the identification of unintended effects. Twenty-two proteins were detected to be differentially expressed in stacked and single GM events compared to a near-isogenic non-GM maize and to a landrace variety [2]. Mesnage et al. generated proteome profiles of Roundup Ready maize and near-isogenic maize varieties using molecular profiling. A comparison of GM maize vs. the near-isogenic maize revealed alterations in the levels of enzymes of the glycolysis and TCA cycle pathways, which can be interpreted as an imbalance in energy metabolism [58]. Benevenuto et al. performed proteomic and metabolomic analyses on NK603 maize plants under abiotic stress conditions. Twenty differentially modulated proteins were identified between GM and non-GM hybrids under water deficiency conditions and herbicide sprays [12]. Barros et al. compared two Roundup Ready maize varieties with a near-isogenic non-GM variety using transcriptomics, proteomics, and metabolomics. The results

showed that the environment had a major effect on protein and gene expression and metabolite production [8]. Harrigan et al. analyzed differences in the metabolome between NK603 hybrids, corresponding negative segregants, and conventional comparator hybrids. It was concluded that the largest effects on metabolomic variation were due to growing location and genomic differences associated with backcrossing practices [37].

In previous studies, we identified a silent mutation in the coding region of the *cry1ab* gene in different stacked MON810 maize varieties using high resolution melting (HRM) analysis, Sanger sequencing, and amplicon sequencing [10, 11]. We have also performed a molecular analysis of NK603 maize and identified two insertions, which are not present in the NK603 patent sequence. These insertions were located in the transgenic promoter region; therefore, they may have an effect on the promoter activity and, consequently, also on transgene expression [14]. However, our analysis of the 5'-end of MON810 as a single event with Scorpion probes revealed no unintended effects or mutations [62].

In this study, we focused on a non-targeted transcriptomics approach (RNA-sequencing) for the identification of possible unintended effects in herbicide-resistant maize (NK603 maize). To the best of our knowledge, this is the first study in which NK603 maize varieties and corresponding non-GM counterparts were analysed by RNA sequencing. We compared the entire transcriptome of RR maize with near-isogenic varieties to characterize gene candidates that may be differentially expressed between RR maize and near-isogenic varieties in two different genetic backgrounds. We conducted a principal component analysis (PCA) as well as a heatmap to evaluate the variance structure of our data. Moreover, we performed GO, KEGG annotations/enrichment and DAVID analyses, which can identify biological functions and metabolic pathways particularly affected by the modification of gene expression. Our results indicate several unintended effects that may have occurred as a result of transgene integration.

## Materials and methods

### Rationale and experimental design

The main objective of this study was to compare gene expression profiles of two different NK603 varieties (AG8025RR and AG9045RR) with the profiles of corresponding near-isogenic varieties that do not contain any transgenes (AG8025conv and AG9045conv, respectively) and investigate whether there are consistent alterations in the NK603 varieties. For this purpose, we used high-throughput RNA sequencing as well as RT-qPCR approach. The pipeline used for the analysis of this study

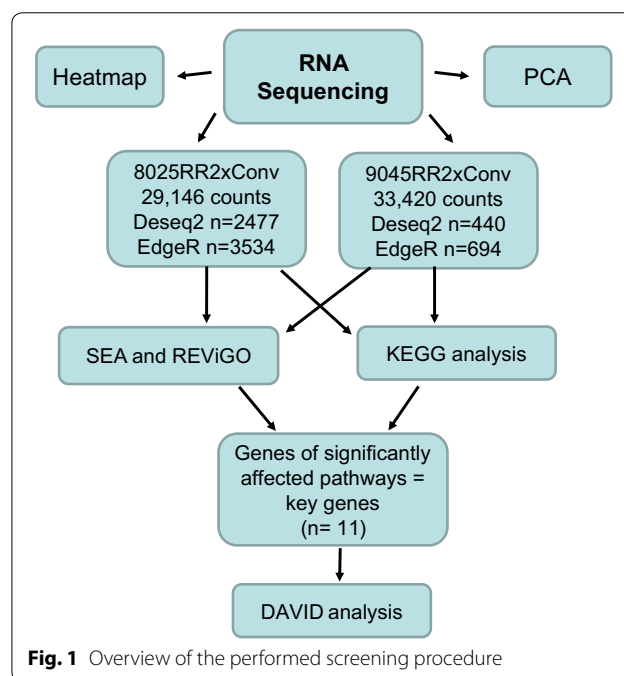
is depicted in Fig. 1. DGE analysis was performed by two different statistical packages (DeSeq2, EdgeR).

Further, we performed a singular enrichment analysis (SEA) to identify if special gene classes and gene interacting networks that are overrepresented among the differentially expressed genes, and we identified the biological functions of differentially expressed genes using GO annotations. In order to search for shared GO terms in the two NK603 genetic backgrounds, we have performed a second joint analysis using REVIGO tool followed by a functional pathway analysis with KEGG annotations. A KEGG metabolic pathways analysis was also conducted. These analyses were performed for each variety pair (NK603 and conventional counterpart) separately. Finally, we compared the genes of the overrepresented gene classes (SEA) with the genes involved in significantly affected metabolic pathways (KEGG) and conducted a DAVID analysis with these key genes to clarify the relationship between these genes.

Next to this, we performed RT-qPCR for determining if some of the detected differences in gene expression of AG8025RR2 vs. AG8025conv were also present in AG9045RR2 vs. AG9045conv. In parallel to these analyses, we have performed a PCA as well as a heatmap to evaluate the variance structure of our data set.

### Plant material

Maize hybrids of the commercial GM varieties AG8025RR2 and AG9045RR2 (unique identifier MON-ØØ6Ø3-6 from Monsanto Company, carrying the NK603



**Fig. 1** Overview of the performed screening procedure

insert which enables glyphosate herbicide tolerance) and the near-isogenic varieties AG8025 (conventional counterpart of AG8025RR2) and AG9045 (conventional counterpart of AG9045RR2) were obtained from the Brazilian market (Sementes Agroceres). AG8025RR2 and its conventional counterpart AG8025 have the same genetic background since they are produced from the same endogamic parental lines. The same is for AG9045RR2 and its conventional counterpart AG9045. The AG8025 variety is the hybrid progeny of the single-cross between maternal endogamous line “A” with the paternal endogamous line “B”. Thus, the tested hybrid variety grains have high genetic similarity (AB genotype). The AG9045 variety is the hybrid progeny of the single-cross between maternal endogamous line “C” with the paternal endogamous line “D” resulting in CD genotype. Presence/absence of the transgenic NK603 insert was confirmed by qPCR in both, GM and conventional maize grains. The obtained maize hybrids were directly used for the transcriptome analysis. We have not received any information about the exact place of cultivation, type of soil, the possible application of fertilizers or pesticides. However, it is confirmed that these varieties were produced in Brazil in 2012.

#### RNA extraction

Total RNA from pools of ten maize grains (of each variety) was isolated based on protocols of Barros et al. and Cheng et al. [8, 15]. Maize grains were homogenized in liquid nitrogen using mortar and pestle. 500 mg of the homogenized grains was transferred into a 15 ml tube and 5 ml of 60 °C warm extraction buffer (2% CTAB, 2% PVP), 2 M NaCl, 0.9 mM DEPC, 0.5 mM spermidine, 100 mM Tris [pH=8]) and 100 µl of mercaptoethanol were added. After vortexing, 5 ml of chloroform-isoamyl-alcohol (24:1; CIA) was added. The mixture was vortexed and centrifuged (15 min, 15 °C, 3184 × g). The upper aqueous phase was mixed with 5 ml of CIA and centrifuged again (15 min, 15 °C, 3184 × g). This CIA extraction was repeated once. After the third centrifugation, 900 µl of 10 M lithium chloride with 1 mM DEPC was added to the upper phase (in a new tube). The mixture was vortexed and stored in the refrigerator (4 °C, o/n). After o/n incubation the mixture was centrifuged (30 min, 5 °C, 3184 × g) and the supernatant was eliminated. The formed pellet was solved with 500 µl of 60 °C warm SSTE buffer [1 M NaCl, 0.5% SDS solution, 1 mM EDTA, 1 mM DEPC, 10 mM Tris (pH=8)]. Then, the buffer/pellet mixture was transferred into a new tube and 500 µl of CIA was added. After vortexing, the mixture was centrifuged (10 min, 21 °C, 16,363 × g). The upper aqueous phase was transferred to a new tube and 96% ethanol was added in the threefold amount of the sample (approx. 1.5 ml). Then, the samples were incubated on ice

(5 min) and centrifuged (4 °C, 35 min, 16,363 × g). After eliminating the supernatant, 250 µl of 75% ethanol were added and the mixture was vortexed shortly. The mixture was centrifuged (4 °C, 10 min, 16,363 × g) and the supernatant was removed. For drying the pellet, the tubes were incubated with open lids (10 min, rt), and 150 µl of 10 mM Tris (pH > 7) was added. After an incubation step (65 °C, 10 min), the RNA was dissolved and extracted RNA was purified (including DNA digestion) using the RNeasy Mini Kit (Qiagen) and the RNase-Free DNase Set (Qiagen) following the manufacturer's instructions. Purified RNA was quantified by a fluorometer (Qubit). The integrity of the purified RNA intended for RNA sequencing was verified with an Agilent 2100 bioanalyzer (Agilent Technologies). RNA was stored at – 80 °C until use.

#### RNA sequencing and data analysis

Three pools of variety AG8025conv and three pools of variety AG8025RR2 were subjected to standard Illumina library preparation using the NEBNext Ultra RNA library prep kit according to the manufacturer's instructions. Three pools of varieties 9045conv and 9045RR2, respectively were treated in the same way. The resulting six cDNA libraries of each conv, RR2 variety pair were paired end sequenced (125 bp) using an Illumina HiSeq2500 machine at the Vienna Biocenter Core Facilities NGS unit (<https://www.vbcf.ac.at>). Reads, which passed basic quality control (Illumina chastity filter), were preprocessed: Adapters were removed with cutadapt (<https://cutadapt.readthedocs.io/en/stable/guide.html>) and the reads were filtered against the rDNA using Bowtie2 (<https://bowtie2-bio.sourceforge.net/bowtie2/index.shtml>), which is a very sensitive contaminants database. The remaining reads were paired end aligned with STAR [25] against the B73 reference genome of *Zea mays* (AGPv3).

#### Statistical analyses

The Principal Component Analysis was performed using ‘stats and ggplot2’ libraries while the heatmap clustering analysis was conducted using ‘pheatmap’ library in R environment aiming to find gene expression patterns across the different varieties. All clean read counts for each sample ( $n=12$ ) were used in these analyses.

In order to determine DEGs, the two NK603 varieties AG9045RR2 and AG8025RR2 were compared to their isogenic counterparts AG9045conv and AG8025conv, respectively. Data were calculated by two tests, DESeq2 and EdgeR, using the software packages Bioconductor and Galaxy. These tests are among the best and most used performance tools for RNA-seq analysis with low numbers of false positives and reliable gene-wise dispersion estimates across all samples [4, 53, 72]. DESeq2 and EdgeR analysis are based on the assumption that the

data follow a negative binomial distribution [27]. Using the raw counts, the data were normalized and transformed to correct for dispersion artifacts and variability within the compared groups or to account for differences in sequencing depths. Genes without any counts were removed. As significance level an adjusted *p*-value controlled for multiple testing using Benjamini and Hochberg's correction- with a false discovery rate (FDR) below or equal to  $\leq 0.05$  was taken for the characterization of DEGs [13, 67, 75, 76]. In addition, a general cut-off threshold of a log<sub>2</sub>-fold change (log<sub>2</sub> FC)  $\geq +1$  as well as a log<sub>2</sub> FC  $\leq -1$  was used. Thus, DEG unigenes had to meet the following criteria: produce a *p*-value below or equal to 0.05 and a log<sub>2</sub>-fold value above log<sub>2</sub> FC  $\geq +1$  for upregulation or produce a *p*  $\leq 0.05$  together with a log<sub>2</sub>-fold value below log<sub>2</sub> FC  $\leq -1$  for downregulation.

DEGs were annotated and calculated for enriched ontologies at a significance level of 5% by AgriGO v.2 [80], a specific GO analysis toolkit and a database for agricultural purposes. There are three ontologies in the GO database, namely, molecular function, cellular component and biological process. For this study, we used the results from the category "biological process". DEGs were also annotated using KEGG pathway enrichment analysis aiming at identifying significantly enriched metabolic pathways or signal transduction pathways affected by the NK603 transgene insertion in each variety. Pathways with adjusted *p*-value  $< 0.05$  were significantly enriched. KEGG analysis was performed using 'clusterprofiler' and 'enrichplot' packages in R environment [86, 87]. For the second KEGG analysis an annotation mapping [56, 85] was performed using Kobas 3.0 (<https://kobas.cbi.pku.edu.cn/kobas3>). KOBAS is a widely used gene set enrichment analysis tool, its annotation module accepts gene list as input, and generates annotations for each gene based on multiple databases about pathways, diseases, and Gene Ontology.

Selected key genes were examined using DAVID Bioinformatics Resources 6.8 (<https://david.ncifcrf.gov/>). After creating a gene list (using Entrez ID), a Gene Functional Classification Analysis was performed. Similarities were measured by Kappa values; furthermore, KEGG pathways were determined. First, a cluster analysis was performed with the following settings: the classification stringency was set to the lowest level, i.e. the individual parameters were as follows: (a) Kappa Similarity: "Similarity Term Overlap level=3", "Similarity Threshold=0.2"; (b) Classification parameters: "Initial Group Membership=3", "Final Group Membership=3", "Multiple Linkage Threshold=3". Subsequently, the similarity between all genes was determined using Kappa scores between 0.05 and 1. Cohen's Kappa values are generally accepted to be a robust measure for genetic similarity.

Kappa results range between 0 and 1. The higher the value of Kappa, the stronger the agreement. Kappa more than 0.7 typically indicates a strong agreement between two genes [57].

Furthermore, KEGG pathways were determined with the genes of the cluster. The components of the pathways found were as follows: "zma-Category; Counts; LT(List-Total); PH (Pop Hits); PT(Pop-Total); %; *p*-Value; Fold Enrichment":

(1) zma00010;3;3;181;6522;100;0,00077;24 / (2) zma01230;3;3;360;6522;100;0,0030;12 / (3) zma01200;3;3;364;6522;100;0,0031;12 / (4) zma01130;3;3;634;6522;100;0,0094;6,9 / (5) zma01110;3;3;1415;6522;100;0,047;3,1 / (6) zma01100;3;3;2496;6522;100;0,15;1,7.

#### Confirmation of differentially expressed genes with RT-qPCR

The gene expression of two differentially regulated genes (GRMZM2G127948, AC204711.3\_FG003) was assayed by RT-qPCR. To test the biological and the technical reproducibility of the RNA-Seq results (validation of RNA-Seq), a new set of three RNA pools was generated from AG8025conv and AG8025RR2 which was not used for the main RNA-Seq assay. In addition, three new pools of AG9045RR2 and its counterpart AG9045conv were generated to confirm the identified differences between the AG8025conv and AG8025RR2 gene expression.

RT-qPCR was performed on a Rotor Gene Q (Qiagen) using the GoTaq<sup>®</sup> 1-Step RT-qPCR System (Promega) according to the manufacturer's instructions. Each reaction was performed in a 20  $\mu$ L final volume containing 10 ng total RNA and 0.2  $\mu$ M of each primer. Primers targeting differentially expressed genes were designed using the software Primer-BLAST (<https://www.ncbi.nlm.nih.gov/tools/primerblast/>). Their sequences are given in the results section. Primer for the reference gene, ubiquitin carrier protein (fwd. 5'-CAGGTGGGGTATTCTTGGTG-3', rev. 5'-ATGTTCCGGGTGGAAAACCTT-3') were taken from Manoli et al. [55]. The primers were purchased from Sigma-Aldrich (Vienna, Austria).

First, to test the specificity of the PCR primer a melting curve analysis was executed which should give just a single peak for each primer pair. Secondly, the efficiency was tested by a dilution series with four different concentrations of a sample. From this test, a standard line and a slope was obtained [68]. Each standard of the dilution was tested in duplicate with a target primer and ubiquitin carrier protein as reference.

All reactions were performed with a hold step at 37 °C (15 min) followed by an initial denaturation step at 95 °C (10 min), followed by 45 cycles of denaturation (95 °C for 10 s), annealing (60 °C for 30 s) and extension (72 °C for 30 s) with a fluorescence measurement at the last step of

each cycle. A melting curve, ranging from 60 to 99 °C, with fluorescence measurements at 1 °C intervals, was done after every RT-qPCR, to determine the specificity of the reaction. The differential expression of the two selected genes was measured by comparing the RNA of three independent pools of GM maize grains with three independent pools of conventional maize grains. Each sample pool was tested in triplicate. One pool consisted of 8 maize grains. For inhibition testing and to evaluate the efficiencies of the specific and reference gene PCR assays, standard curves were pipetted in every run. Negative-control and reverse transcriptase (RT)-minus controls (reverse transcription reaction without addition of reverse transcriptase) were also used. Every run was repeated on another day and the mean of the values was taken.

Data were evaluated using the Rotor Gene Q Series software. Linearity (R2) and efficiency ( $E = 10^{[-1/\text{slope}]}$ ) of every reaction were within the accepted values. For a valid linearity a value of >0.98 and for a valid efficiency a slope between -3.1 and -3.6 was required. Relative quantitation was calculated using the  $2^{-\Delta\Delta C_t}$  method [52]. Values were normalized using the reference gene ubiquitin and efficiency correction was performed as described by Pfaffl [64].

## Results

### 1\_RNA sequencing (RNA-seq) and library construction for NK603 and near-isogenic maize kernels

Three cDNA libraries per variety were constructed for each GM variety and their near-isogenic non-GM counterparts (12 libraries in total). Table 1 gives an overview of the statistical analyses and the bioinformatic data processing of the libraries per variety. The libraries of AG8025RR2 and AG8025conv yielded 93.6 M of raw reads on average. A total of 14.6% of the raw reads were removed due to adapter removal and cleaning (quality filtering and elimination of rRNA). Next, 73.8% of the raw reads could be mapped uniquely to the B73 maize reference genome version AGPv3. Overall, 5.7% of the raw reads contained repetitive sequences, which aligned at multiple sites. Further, 5.9% of the reads could not be

mapped because they had too many errors to match the target sequence. Uniquely mapped reads corresponded to 39,625 unigenes containing 29,146 genes with counts above zero in the transgenic and near-isogenic varieties. The libraries of AG9045RR2 and AG9045conv yielded 71.9 M of raw reads on average; 9.2% of these were removed, 48.6% could be mapped, 4.6% contained repetitive sequences, and 16.4% of the reads could not be mapped. Uniquely mapped reads corresponded to 39,625 unigenes containing 33,420 genes with counts above zero in the transgenic and near-isogenic varieties.

### 2\_Exploratory data—PCA and heatmap

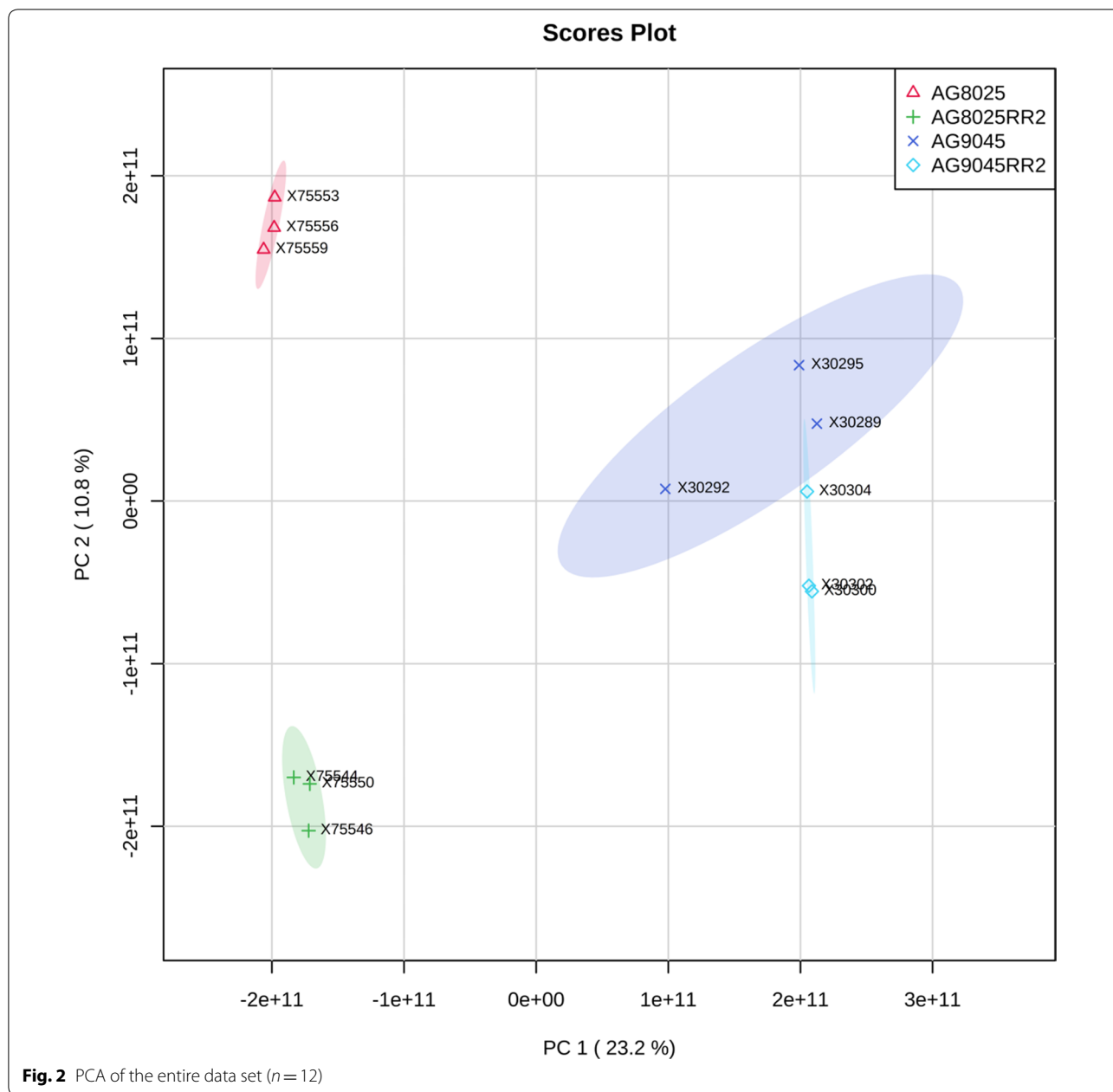
Unsupervised PCA results are depicted in Fig. 2 for the entire data set. The analysis demonstrated a clear cluster by variety (PC1 23% of variation) and a second cluster by transgenic event (PC2 11% of the variation). Variability within the replicates was low (except for one sample from AG9045). For AG9045 and AG9045RR2 the distance between transgenic and non-transgenic is lower than the distance between both non-transgenic. For AG8025 and AG8025RR2 the distance between transgenic and non-transgenic is similar to the distance between both non-transgenic. The heatmap data are presented in Fig. 3. The heatmap hierarchical clustering corroborates the PCA results as the varieties pairs are clustered together. In addition, there was no significant variability between replicates observed.

### 3\_Characterization of significant differentially expressed genes

Based on the criteria outlined in section [Materials and methods](#) (subchapter statistical analyses) a total of 2477 genes were determined to be differentially expressed by DESeq2 in AG8025RR2 vs. AG8025conv. (29,146 unigene counts above zero), and a total of 440 genes were determined to be differentially expressed in AG9045RR2 vs. AG9045conv. (33,420 unigene counts above zero). EdgeR analysis indicated 3534 and 694 DEGs were present in AG8025RR2 vs. AG8025conv and in AG9045RR2 vs. AG9045conv, respectively. The DEGs and analytical process results are shown in Fig. 1. The complete list of

**Table 1** Details of run statistics and data processing for library construction AG8025RR2, AG9045RR2 and near isogenic varieties

Run statistics	AG8025RR2	AG8025 conv	AG9045RR2	AG9045 conv
Raw reads	95.5 M	91.7 M	71.1 M	73.1 M
Uniquely mapped reads	71.7 M	65.9 M	37.6 M	49 M
Repetitive sequences	5.4 M	5.2 M	3.1 M	3.6 M
Not aligned	5.0 M	6.1 M	20.8 M	2.8 M
Median insert size (nt) (from aligned paired end reads)	107	128	165	159



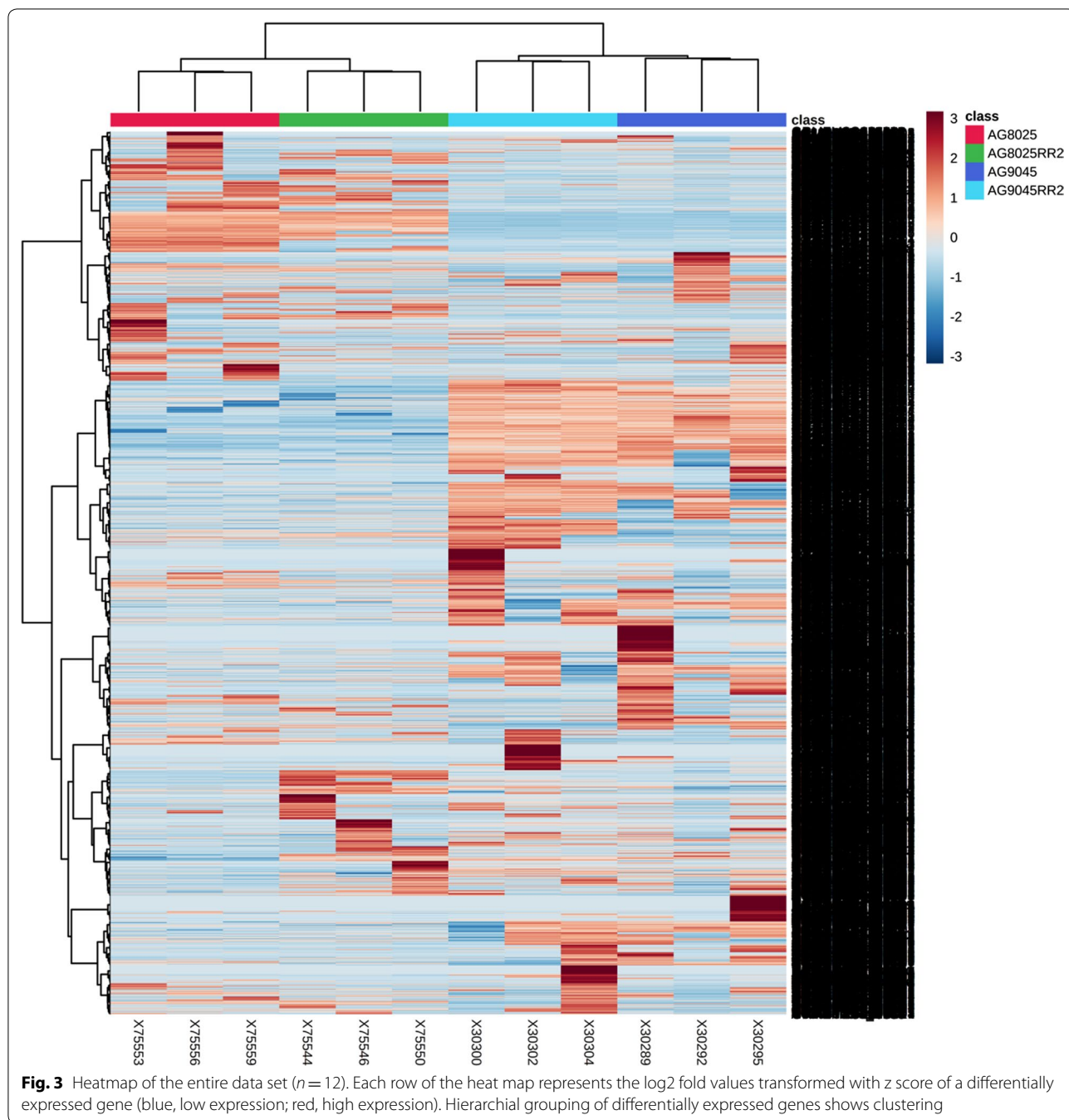
all DEGs analyzed by both tools are available under Additional files: 1–8.

Analysis of the data using EdgeR and DESeq2 showed that a total of 1355 genes were significantly upregulated and 1105 genes were significantly downregulated in the AG8025RR2 variety vs. the AG8025conv counterpart. On the other hand, in the AG9045 variety, a total of 308 genes were significantly upregulated, and 79 genes were significantly downregulated in the AG9045RR2 variety vs. the AG9045conv counterpart. If just common genes were considered, 27 genes were commonly upregulated,

and 26 common genes were downregulated in both varieties when comparing the significantly upregulated or downregulated genes of the two varieties AG8025 and AG9045.

#### 4\_Singular enrichment analysis, characterization of Gene Ontologies, and REVIGO

After statistically defining DEGs, these genes were further analyzed by SEA using the agriGO tool. Exploring GO has become a widespread practice to obtain insights into the potential biological meaning of RNA



experiments. For this analysis, there is a growing number of available gene-sets and functional literature containing many genes/proteins and biochemical pathways. Computational analysis helps to find functionally coherent gene-sets that are statistically overrepresented in a given gene list. The results are mapped as well and compared to gene functional categories of a GO database. Thus, GO analysis may indicate that a certain biological process plays a role in the analyzed biological condition.

For the SEA analysis, the sum of the upregulated plus downregulated DEGs of both varieties, AG8025RR2 vs. AG8025 conv ( $n=2460$ ) and AG9045RR2 vs. AG9045conv ( $n=387$ ), was calculated (Additional files 9, 10). There were  $n=81$  GOs in the AG8025RR2 vs. AG8025conv comparison and  $n=78$ -GOs in the AG9045RR2 vs. AG9045conv comparison. However, these numbers were reduced when considering just the



**Table 2 Characteristics of significant GO-ontology groups before (all rows) and after REViGO calculation (marked in bold)**

GO:ID	Description	No. genes AG8025 vs AG8025RR2	No. genes AG9045 AG9045RR2	No. of common genes	SEA_FDR mv	log10_p-val	Frequency (%)	Uniquen	Dispens
GO:0003006	development process involved in reproduction	350	71	22	<b>0.0310</b>	-1.5086	1.227	<b>0.92</b>	<b>0.00</b>
GO:0042221	Response to chemical	726	150	42	<b>0.0000</b>	-4.4909	4.749	<b>0.34</b>	<b>0.00</b>
GO:0044711	Single-organism biosynthetic process	534	91	24	<b>0.0160</b>	-1.7959	6.613	<b>0.87</b>	<b>0.00</b>
GO:0044763	Single-organism cellular process	1398	238	67	<b>0.0177</b>	-1.7533	19.396	<b>0.87</b>	<b>0.3</b>
GO:0009651	Response to salt stress	178	40	14	<b>0.0024</b>	-2.6289	0.389	<b>0.37</b>	<b>0.41</b>
GO:0009266	Response to temperature stimulus	167	43	10	0.0020	<b>-2.6946</b>	0.601	0.37	0.74
GO:0006970	Response to osmotic stress	186	46	15	0.0016	<b>-2.7945</b>	0.468	0.36	0.71
GO:0044710	Single-organism metabolic process	1112	181	51	<b>0.0210</b>	-1.6778	<b>23.696</b>	<b>0.87</b>	<b>0.42</b>
GO:0051707	Response to other organism	293	63	17	<b>0.0014</b>	-2.8697	0.704	<b>0.33</b>	<b>0.44</b>
GO:0043207	Response to external biotic stimulus	294	64	17	0.0014	<b>-2.8697</b>	0.704	0.33	0.97
GO:0009620	Response to fungus	120	26	9	0.0137	<b>-1.8649</b>	0.285	0.36	0.91
GO:0098542	Defense response to other organism	120	26	9	0.0083	<b>-2.0835</b>	0.601	0.31	0.98
GO:0009607	Response to biotic stimulus	311	66	17	<b>0.0019</b>	-2.7328	0.771	<b>0.42</b>	<b>0.44</b>
GO:0009605	Response to external stimulus	423	82	24	<b>0.0011</b>	-2.9788	1.057	<b>0.41</b>	<b>0.46</b>
GO:1901700	Response to oxygen-containing compound	428	101	29	<b>0.0006</b>	-3.2217	1.196	<b>0.35</b>	<b>0.47</b>
GO:0010033	Response to organic substance	519	121	36	0.0010	<b>-3.0222</b>	2.423	0.31	0.71
GO:0070887	Cellular response to chemical stimulus	349	76	16	0.0186	<b>-1.7317</b>	1.470	0.34	0.73
GO:0006952	Defense response	330	64	15	0.0036	<b>-2.4461</b>	1.312	0.37	0.48

**Table 2 (continued)**

GO:ID	Description	No. genes AG8025 vs AG8025RR2	No. genes AG9045 AG9045RR2	No. of common genes	SEA_FDR mv	log10_p-val	Frequency (%)	Uniquen	Dispens
GO:0009628	Response to abiotic stimulus	462	109	25	0.0004	-3.4144	1.925	0.38	0.5
GO:0009719	Response to endogenous stimulus	388	95	27	0.0125	-1.9031	2.168	0.38	0.51
GO:0010035	Response to inorganic substance	261	52	14	0.0108	-1.9666	0.735	0.37	0.62
GO:0006950	Response to stress	821	146	39	0.0077	-2.1135	7.803	0.31	0.62
GO:0001101	Response to acid chemical	290	71	19	0.0075	-2.1240	0.753	0.37	0.62
GO:0009409	Response to cold	125	30	8	0.0031	-2.5157	0.310	0.38	0.68

GO:ID, identity of GO-Ontologies; description, definition of ontologies; nr.genes AG8025 vs AG8025RR2, and nr.genes AG9045 vs AG9045RR2, number of genes in the ontology groups AG8025 vs AG8025RR2, and AG9045 vs AG9045RR2; common genes(nr.), number of genes commonly up- or downregulated genes in both variety pairs; SEA\_FDR mv, mean value of AG8025 and AG9045 false discovery rate ( $p < 0.05$ ); Rev log10, log10 value of REViGO analysis; Frequ, proportion of GO terms in the annotation database; Uniqueness, comparison measure to other ontologies (Parameter for ontology outliers); Dispensability, similarity threshold for elimination of ontologies

common ontologies of both varieties which resulted in 24 GO ontologies (Table 2).

Redundancy of GM terms constitutes a major problem for the interpretation of RNAseq results. Very recently, a software called REViGO was developed using an algorithm that evaluates semantic similarities in GO assignments from hierarchical functional annotation of gene ontologies [79]. GO terms with semantic similarities are used to find representative subsets of the terms. Clusters are formed, and each of the GO terms is assigned to the clusters. Cluster representatives are kept, while subordinate cluster members can be removed, thus avoiding redundancy in the results. However, it should be considered that GO annotations are hierarchical, REViGO retains more general ontologies and removes more detailed ontologies, and other programs e.g. ClueGO sort ontologies differently. The usage of REViGO enabled us to reduce the number of ontologies from 24 to 16 GOs. When taking a closer look at the total amount of genes, we discovered  $n=81$  unique genes (see Table 3) within the 16 ontologies. The results, in particular the number of genes in the GO clusters, the FDR values from the SEA analysis, as well as the p-values for the frequency, uniqueness and dispensability—these parameters are useful to select different level of ontologies—of the SEA analysis are given in Table 2. The largest groups of ontologies were different types of responses, for instance, response to stress, cold, inorganic substances, chemicals, acid chemicals, oxygen-containing compounds, and salt

stress, the defense response, and the response to abiotic, biotic, chemical, external or endogenous stimuli. Next, a smaller group of ontologies contained three single-organism processes: a biosynthetic process, single-organism metabolic process and single-organism cellular process. Other ontologies were developmental processes involved in reproduction and single-organism biosynthetic processes.

### 5\_KEGG pathway classification of identified genes

The GO ontologies evaluated in the previous section contain just a general semantic description of DEGs, which are merged into certain GO terms. One general drawback of GO terms is that their meaning may be very general. However, an important aspect of this study should be that the biological/biochemical functions of identified genes and their functional annotation are understood. This is possible by means of pathway analysis using KEGG annotation/enrichment analysis of single DEGs. KEGG is a database containing a collection of genomes, biological pathways, diseases, drugs, and chemical substances. Thus, KEGG is a popular method to find functionally related genes and pathways that are enriched in a gene list and can be defined based on participation in a metabolic or signaling pathway.

For the KEGG analysis, the two NK603 varieties, AG9045RR2 and AG8025RR2, were compared to their isogenic counterparts, AG9045conv and AG8025conv, respectively. First, an annotation mapping was performed

**Table 3 Gen-IDs, occurrence in GO-Ontologies, ko-identities and gene description of GO-Ontology genes**

Gene-ID	Occur. in 16 GOs	Entrez-ID	Gene/protein description
AC204711.3_FG003	8	100286331	Senescence-associated protein DIN1
GRMZM2G005954	16	100191257	Uncharacterized
GRMZM2G007252	4	100383338	3beta-hydroxysteroid-dehydrogenase/ Decarboxylase isoform 2
GRMZM2G011523	1	100276951	Uncharacterized
GRMZM2G013448	11	100283053	1-Aminocyclopropane-1-carboxylate oxidase 1
GRMZM2G025611	5	103629239	Cation/H(+) antiporter 20
GRMZM2G028677	1	100284998	Trans-cinnamate 4-monooxygenase
GRMZM2G036351	6	100284433	Uncharacterized
GRMZM2G039166	9	541876	Homocysteine S-methyltransferase 4
GRMZM2G040559	1	100281010	Tubulin alpha-6 chain
GRMZM2G046392	1	Does not match to Entrez	GenBank accession: AFW62113
GRMZM2G053206	4	100272880	Uncharacterized
GRMZM2G054224	9	100274058	Putative prolyl 4-hydroxylase 12
GRMZM2G069758	11	100283562	Pyridoxal kinase
GRMZM2G084491	8	103641981	Uncharacterized
GRMZM2G089557	16	100281848	Zinc finger, C2H2 type family protein
GRMZM2G090043	1	100275716	Uncharacterized
GRMZM2G099619	10	100501607	ABC transporter G family member 16
GRMZM2G105987	10	100281231	NHL25
GRMZM2G106445	2	100285556	Wound induced protein
GRMZM2G107073	2	100284797	Glycoside hydrolase, family 28
GRMZM2G127251	2	100274269	Hydroxycinnamoyltransferase3
GRMZM2G127948	3	107282093	Caffeoyl-CoA O-methyltransferase1
GRMZM2G129018	16	100284454	Uncharacterized
GRMZM2G129674	5	100216626	Uncharacterized
GRMZM2G135044	6	100272362	Uncharacterized
GRMZM2G144997	1	103634596	Chaperone protein dnaJ 20, chloroplastic
GRMZM2G148485	2	103653873	Expansin-B15
GRMZM2G152175	3	100284504	Dihydroflavonol-4-reductase
GRMZM2G156356	6	100283084	Maltose excess protein 1-like
GRMZM2G156632	1	542729	Wound induced protein 1
GRMZM2G162359	12	100274481	Uncharacterized
GRMZM2G166944	10	100191584	Xyloglucan endotransglucosylase/hydrolase protein 23
GRMZM2G306643	4	100278004	Senescence regulator
GRMZM2G354319	6	103629679	Putative B-box type zinc finger family protein
GRMZM2G408158	11	103648268	9-cis-epoxycarotenoid dioxygenase NCED4
GRMZM2G425206	2	100280859	OLE-5
GRMZM2G446253	2	103631112	Enolase 1-like
GRMZM2G477872	14	100280163	Uncharacterized
GRMZM5G800586	10	100273184	BURP domain protein RD22
GRMZM5G831577	11	100280792	RNA-binding protein
GRMZM5G880028	4	103639108	Putative calmodulin-binding family protein
GRMZM5G896756	1	Does not match to Entrez	GenBank accession: AFW74673
GRMZM5G899851	5	542229	Uncharacterized
GRMZM2G003711	1	Does not match to Entrez	GenBank accession:DAA60655
GRMZM2G015599	1	100272609	Uncharacterized
GRMZM2G018059	7	103646429	Uncharacterized
GRMZM2G020423	1	100280615	Uncharacterized

**Table 3 (continued)**

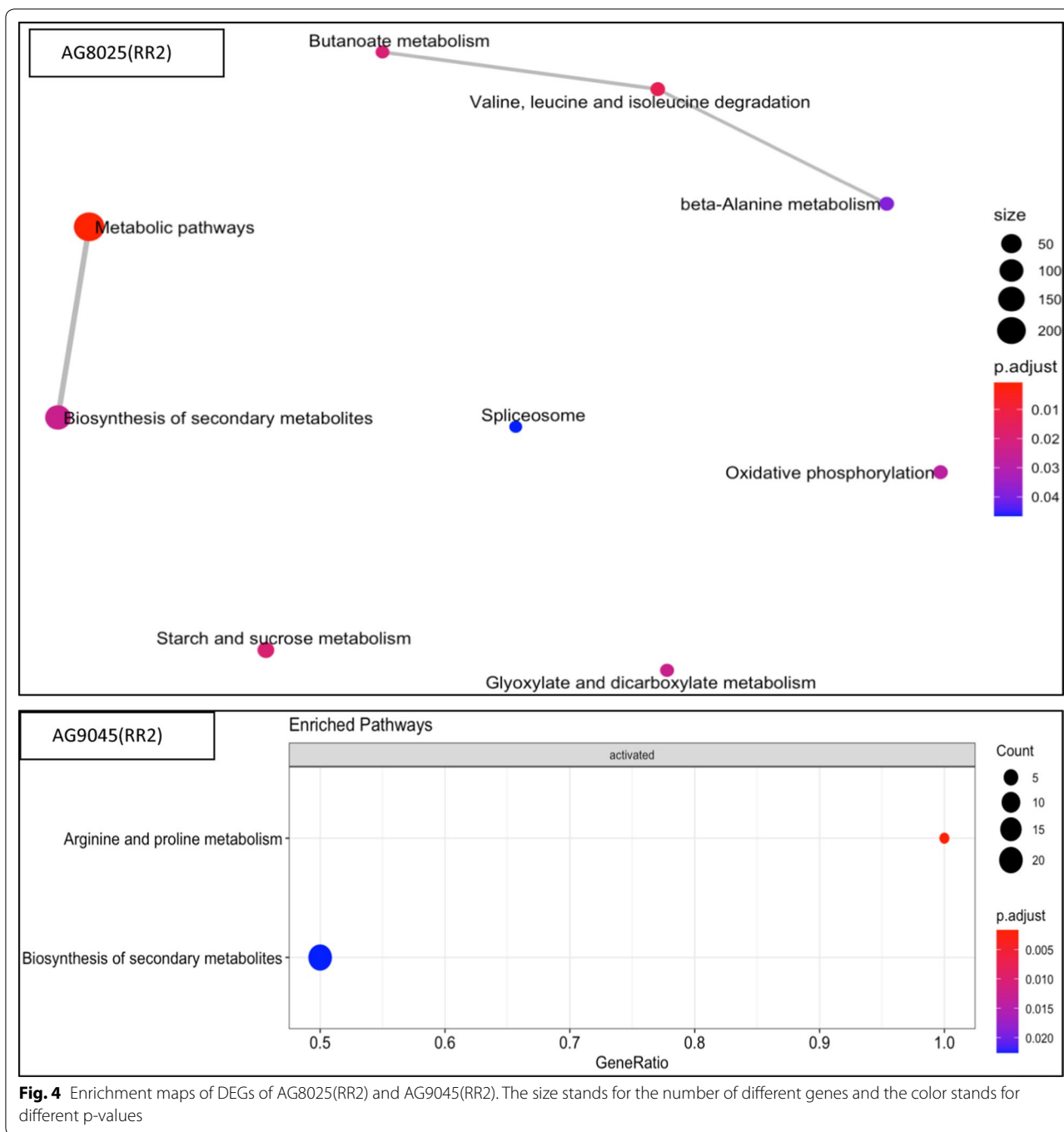
Gene-ID	Occur. in 16 GOs	Entrez-ID	Gene/protein description
GRMZM2G027490	1	Does not match to Entrez	GenBank accession:DAA63464
GRMZM2G028521	2	100285151	Uncharacterized
GRMZM2G044670	2	100272471	Uncharacterized
GRMZM2G071630	12	542333	Cytosolic glyceroldehyde-3-phosphate dehydrogenase
GRMZM2G075018	1	100383107	Uncharacterized
GRMZM2G087714	2	107198085	Protein NUCLEAR FUSION DEFECTIVE 6 Chloroplastic/mitochondrial
GRMZM2G089750	5	100193060	Uncharacterized
GRMZM2G103342	14	100275716	Uncharacterized
GRMZM2G104942	2	100216884	Putative glycerol-3-phosphate transporter 1
GRMZM2G105657	1	100274546	Uncharacterized
GRMZM2G106092	3	Does not match to Entrez	GenBank accession:DAA38807
GRMZM2G108847	1	100037813	Putative serine type endopeptidase inhibitor
GRMZM2G108991	4	103647628	Serine/threonine protein phosphatase 2A 57 kDa regulatory subunit B' alpha isoform
GRMZM2G111189	1	100383286	Uncharacterized
GRMZM2G135536	9	100285285	Uncharacterized
GRMZM2G149422	4	103654240	Protein EXORDIUM
GRMZM2G152908	8	542247	Sucrose synthase 1
GRMZM2G168115	3	100281436	Uncharacterized
GRMZM2G170017	6	100281076	Uncharacterized
GRMZM2G176225	12	100281741	Uncharacterized
GRMZM2G176307	12	100037774	Uncharacterized
GRMZM2G179138	2	Does not match to Entrez	GenBank accession:DAA47206
GRMZM2G180246	1	100101544	Growth-regulating-factor-interacting factor1
GRMZM2G312110	8	103633920	Auxin transport protein BIG
GRMZM2G339336	2	Does not match to Entrez	GenBank accession: AFW64680
GRMZM2G368886	1	107325939	Alpha-expansin4
GRMZM2G373124	8	100280958	Uncharacterized
GRMZM2G703490	1	100278319	Uncharacterized
GRMZM5G801436	3	100282335	Uncharacterized
GRMZM5G817651	7	100278060	Uncharacterized
GRMZM5G836174	5	100191227	Phosphatase phospho 1
GRMZM5G843555	9	100272381	Uncharacterized
GRMZM5G845532	2	Does not match to Entrez	GenBank accession: AFW62896)

using Kobas 3.0. The mapping was carried out with the unique  $n=81$  genes (73 Entrez-IDs) of the 16 ontology groups (these genes are listed in Table 3). The detailed result of the annotations is specified in Additional file 11. We have also performed a KEGG pathways scatter plot for the enriched metabolic pathways for each of the NK603 varieties. Biosynthesis of secondary metabolites was found altered in both NK603 varieties compared to their near-isogenic conventional counterpart. The variety AG8025RR showed imbalance for starch and sugar metabolism, oxidative phosphorylation, glyoxylate and dicarboxylate metabolism, butanoate, beta-alanine,

valine, leucine and isoleucine metabolism and also spliceosome (Fig. 4). Whereas variety AG9045RR only showed arginine and proline metabolism disturbance (Fig. 4). The list of enriched pathways and their corresponding  $p$  adjusted values are present in Table 4 and in Additional file 12.

#### 6\_DAVID—functional classification of genes

We compared the genes of the overrepresented gene classes (SEA) with the genes involved in significantly affected metabolic pathways (KEGG). Genes occurring in SEA as well as in KEGG analysis of AG8025(RR2)



and AG9045(RR2) are depicted in Table 5. If the genes of Table 5 are to be used for more detailed analysis, the knowledge of the genetic similarity between these genes would be very valuable. Thus, the relatedness between the most important genes of Table 5 was determined. Eleven genes occurring in the GO-ontologies as well as in the KEGG pathways of AG8025 and AG9045 were analyzed using the gene functional classification tool

of the DAVID Bioinformatics Package (see [Materials and methods](#)). The analysis revealed a cluster of three genes (Entrez No. 103631112, 100037774, 542333), while eight genes of the list were not in the output. The genetic similarity between all eleven genes was determined by the calculation of Kappa scores (see Table 6). In addition to the determination of similarities, the three cluster genes were also used to determine

**Table 4 Main values of KEGG enrichment for AG8025(RR2) and AG9045(RR2) groups**

ID	Description	Size	Enrich.	NES	p value	p. adjust	q values	Rank
<i>KEGG enrichment 8025</i>								
zma01100	Metabolic pathways	253	-0.176	-2047053	0.0009	0.0009	0.0741	2039
zma00280	Valine, leucine and isoleucine degradation	7	-0.604	-1854659	0.0135	0.0135	0.3370	986
zma00500	Starch and sucrose metabolism	21	0.380	1739141	0.0192	0.0192	0.3370	1008
zma00650	Butanoate metabolism	5	-0.667	-1753212	0.0197	0.0197	0.3370	187
zma01110	Biosynthesis of secondary metabolites	140	-0.156	-1555442	0.0244	0.0244	0.3370	2028
zma00630	Glyoxylate and dicarboxylate metabolism	9	-0.505	-1739344	0.0244	0.0244	0.3370	68
zma00190	Oxidative phosphorylation	7	-0.560	-1721144	0.0277	0.0277	0.3370	720
zma00410	Beta-Alanine metabolism	7	-0.542	-1662939	0.0389	0.0389	0.4152	862
zma03040	Spliceosome	4	-0.660	-1567797	0.0466	0.0466	0.4266	479
<i>KEGG enrichment 9045</i>								
zma00330	Arginine and proline metabolism	3	0.892	1822906	0.0017	0.0017	0.0323	45
zma01110	Biosynthesis of secondary metabolites	40	0.323	1714550	0.0226	0.0226	0.2139	117

ID identification of the metabolic pathways under KEGG database, size the number of genes in that pathway, enrich the enrichment score and NES the normalized enrichment score

**Table 5 Key genes**

Gene-ID	Entrez-ID	Gene/protein description
GRMZM2G007252	100383338	3Beta-hydroxysteroid-dehydrogenase/
GRMZM2G013448	100283053	1-Aminocyclopropane-1-carboxylate oxidase 1
GRMZM2G028677	100284998	Trans-cinnamate 4-monooxygenase
GRMZM2G039166	541876	Homocysteine S-methyltransferase 4
GRMZM2G054224	100274058	Putative prolyl 4-hydroxylase 12
GRMZM2G408158	103648268	9-Cis-epoxycarotenoid dioxygenase NCED4
GRMZM2G446253	103631112	Enolase 1-like
GRMZM2G071630	542333	Cytosolic glyceraldehyde-3-phosphate dehydrogenase
GRMZM2G170017	100281076	Uncharacterized
GRMZM2G176307	100037774	Uncharacterized
GRMZM5G843555	100272381	Uncharacterized

Only genes involved in significantly affected metabolic pathways (KEGG) and among overrepresented gene classes (SEA) are shown

enriched KEGG pathways by DAVID. The following 6 KEGG pathways were found: zma00010: Glycolysis / Gluconeogenesis, zma01230: Biosynthesis of amino acids, zma01200: Carbon metabolism, zma01130: Biosynthesis of antibiotics, zma01110: Biosynthesis of secondary metabolites, zma01100: Metabolic pathways. Two of these pathways (zma01110 and zma01100) were also found with the above-mentioned KEGG enrichments of all significant DEG genes. Thus, a more accurate representation of the components of these KEGG pathways is intriguing. However, since the KEGG pathways contain a huge number of genes, the number of the background genes occurring in the maize database and the complexity of the maps of KEGG pathways were too high to be listed in the results section, but are indicated in Additional files 14 and 15 instead. Based on this information, it is possible to look up

the biochemical relationships between the 11 genes of Table 5.

### 7\_Verification of RNA sequencing results by RT-qPCR

As described above, two different program packages were used for analysis of the primary RNA-seq data with the aim of avoiding false-positive results. However, as in most transcriptomic studies our RNA-seq data contain just a low number of biological replicates ( $n=3$ ), but a high number of total data. Thus, from a mathematical perspective it is desirable, and it is practiced as a gold standard to validate DEG data with additional methods such as microarray hybridization or RT-qPCR analysis [18, 22, 44, 49, 69, 89]. Reverse transcription followed by PCR is a powerful tool for the quantification and detection of gene expression levels, in particular for low-abundance transcripts. Thus, we decided to assign

**Table 6 Functional classification of the key genes indicated by Kappa\* scores**

Gene identity	Gene description	100283053	541876	100284998	103648268	103631112	100383338	100037774	100274058	100281076	542333	100272381
100283053	1-aminocyclopropane-1-carboxylate oxidase 1(LOC100283053)	Kappa = 1,0										
541876	homocysteine S-methyltransferase 4(hmt4)	0.34	Kappa = 1,0									
100284998	trans-cinnamate 4-monoxygenase(LOC100284998)	0.3	0.16	Kappa = 1,0								
103648268	9-cis-epoxycarotenoid dioxygenase 1, chloroplastic-like(LOC103648268)	0.24	0.18	0.15	Kappa = 1,0							
103631112	enolase 1-like(LOC103631112)	0.21	0.17	0.14	0.33	Kappa = 1,0						
100383338	uncharacterized LOC100383338(LOC100383338)	0.2	0.08	0.14	0.15	0.27	Kappa = 1,0					
100037774	uncharacterized LOC100037774(TIDP3317)	0.18	0.1	0.14	0.15	0.41	0.27	Kappa = 1,0				
100274058	uncharacterized LOC100274058(LOC100274058)	0.17	0.07	0.19	0.12	0.11	0.11	0.06	Kappa = 1,0			
100281076	uncharacterized LOC100281076(umc2600)	0.12	0.1	0.08	0.22	0.18	0.17	0.15	<0.05	Kappa = 1,0		
542333	cytosolic glyceroldehyde-3-phosphate dehydrogenase GAPC3(GAPC3)	0.11	0.07	0.1	0.09	0.25	0.16	0.7	<0.05	0.09	Kappa = 1,0	
100272381	uncharacterized LOC100272381(LOC100272381)	0.09	<0.05	0.12	<0.05	<0.05	<0.05	<0.05	0.64	<0.05	<0.05	Kappa = 1,0

two randomly selected differentially regulated genes for reverse transcription real-time PCR. One of the genes was upregulated (AC204711.3\_FG003), while the other was downregulated (GRMZM2G127948). The results of the DEGs (log2 FC values) tested with DeSeq2 and EdgeR for both genes are depicted in Table 7. The log2 FC values for these genes were either above the cut-off value of log2 FC = +1 (upregulated) or below log2 FC = -1 (downregulated).

The first gene, GRMZM2G127948, codes for the protein 'Caffeoyl-CoA O-methyltransferase 1'. Caffeoyl-CoA omethyltransferase 1 is a key enzyme in lignin biosynthesis. Lignin provides mechanical strength to vascular tissues and protects plants from biotic stresses, including pathogen attack [83]. In investigating the ontologies to which these genes belong, the gene with the ID GRMZM2G127948 could be found in the following ontologies of both varieties—AG9045 as well as AG8025:

GO:0044710 (single-organism metabolic process),

GO:0044711 (single-organism biosynthetic process), and

GO:0044763 (single-organism cellular process).

The second gene with the ID AC204711.3\_FG003 is a gene (senescence-associated protein DIN) relevant for senescence-related processes. Senescence is involved in the deterioration found in several parts of a plant or functional characteristics at a cellular level, in death rates or in fecundity. AC204711.3\_FG003 can be assigned consistently to the following functions in the comparison of both varieties—AG9045NK603 vs. AG9045 and AG8025NK603 vs. AG8025:

GO:0044710 (single-organism metabolic process),

GO:0044711 (single-organism biosynthetic process),

GO:0044763 (single-organism cellular process),

GO:0009719 (response to endogenous stimuli),

GO:0042221 (response to chemicals),

GO:0001101 (response to acid chemicals),

**Table 7 Comparison of log2 FC-values of two single genes (AC204711.3\_FG003, GRMZM2G127948) obtained by DGE (DESeq2, EdgeR) and single gene (RT-qPCR) analysis**

AC204711.3_FG003 Senescence-associated protein DIN1		
DGE analysis	AG8025	AG9045
DeSeq2	+ 2.20	+ 1.34
EdgeR	+ 2.92	+ 2.20
RT-qPCR analysis	+ 5.08	+ 2.32
GRMZM2G127948 Caffeoyl-CoA O-methyltransferase1		
DGE analysis	AG8025	AG9045
DeSeq2	-1.81	-1.29
EdgeR	-1.97	-1.45
RT-qPCR analysis	-1.21	-0.82

GO:0006950 (response to stress), and.

GO:1901700 (response to oxygen-containing compounds).

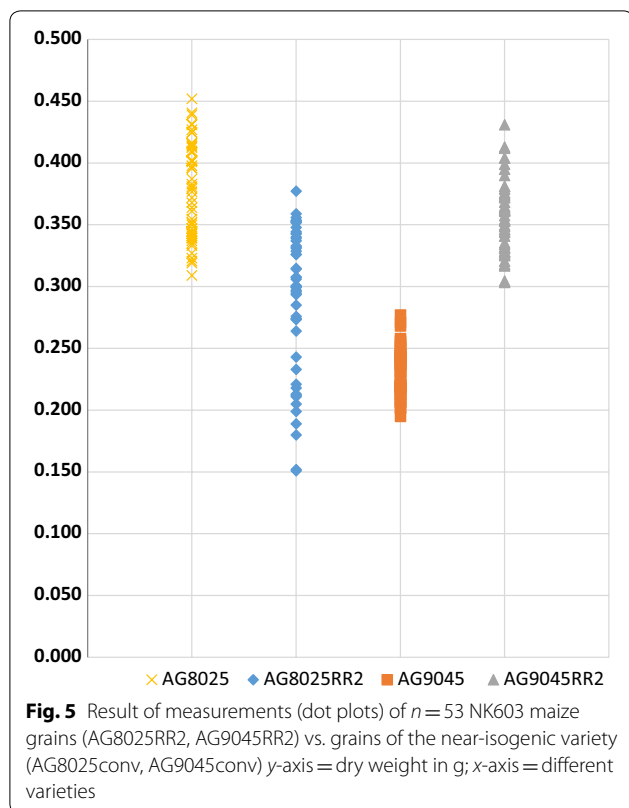
To assess if the DGE analysis of both genes can be validated with independent methods, these genes were tested twice using RT-qPCR with the aid of gene-specific and reference primer pairs (see [Materials and methods](#) section). The log2 FC data obtained with RT-qPCR compared to the RNA-seq data are shown in Table 7. The log2 FC values of the upregulated gene were distinctly above the positive cutoff value in both GM varieties, log2 FC = +5.08 and +2.20 respectively. Additionally, the tendency of the downregulated gene was in accordance with the DGE data. The log2 FC value of the first variety (log2 FC = -1.21) was below the cut-off value (log2 FC = 1), and the log2 FC value of the second variety (log2 FC = -0.82) was very close to the negative cut-off value (log2 FC = -1). For these two genes, the log2 FC values measured by RT-qPCR expression were in good agreement with the DGE data. Thus, the RNA-seq results agreed with the RT-qPCR results and seem to be of high reliability. Nonetheless, it would be still necessary to conduct more single gene experiments using RT-qPCR as a method to get a better estimate of the reliability of the implemented RNA-seq results.

**7\_Measurement of NK603 grains vs. grains of the near-isogenic conventional variety**

When comparing gene expression of GM varieties with those of conventional counterpart plants, it should be considered that potential nonspecific effects might influence the results of the RNA-seq analysis. Thus, we were interested in whether a general difference exists among the investigated kernels. To get a rough estimate of potentially unspecific differences, we measured the weight of GM and conventional kernels. To determine the differences between transgenic and conventional maize, 53 grains of each variety were weighed, and evaluated for significant differences. Because the weights of the grains were not normally distributed (8025conv  $p=0.031$ ; 8025RR2  $p=0.00074$ ; Shapiro Wilk test) we evaluated the significance of the differences by a non-parametric test (Mann Whitney  $U$ -test). AG8025RR2 maize grains had an average weight of 0.291 g and AG8025conv maize grains 0.378 g. This difference was significantly different ( $z$ -value = +7.12, limit for  $p=0.001$  significance = +3.29). Additionally, AG9045RR2 maize grains had an average weight of 0.357 g and AG9045conv maize grains an average weight of 0.232 g. This difference was significantly different as well ( $z$ -value = -8.87, limit for  $p=0.001$  significance = -3.2).

Thus, in the case of the AG8025 comparison, the conventional variety was heavier, whereas in the case of the





AG9045 comparison, the transgenic variety was heavier. Figure 5 shows the dry weights of the maize grains of all tested varieties. All measurement values for these samples are present in Additional file 13. As the results were opposite in the two varieties, it cannot be concluded that there is a consistent weight difference between transgenic maize and its conventional counterpart.

## Discussion

### Rationale of the study

Gene variation and interactions are common and important phenomena in understanding plant genetics and breeding. Thus, a high-throughput RNA sequencing approach allows a dynamic and functional analysis of maize genetics. The development of ‘omics’ technology has enabled comprehensive analysis of gene interactions, as well as unintended effects, in different GM events on transcript, protein and metabolite levels [23, 27]. However, due to differences in methodological approaches and/or genetic background, little to no consistent results have been obtained among previous studies on how gene expression is influenced by transgene insertion and expression in plant genomes.

The rationale of this study was to investigate potential unintended effects deriving from the insertion of a specific transgene—NK603—into two transgenic maize

varieties. RNA-seq analysis was used as a molecular profiling technique to study two GM crops in comparison to their near-isogenic maize varieties. Differentially expressed genes were detected with high stringency, taking into account different varieties. False-positives were limited, as we employed different statistical packages for DGE analysis [4, 16, 53]. Overall, this approach allowed the detection of common effects in two variety pairs. Between the variety pair AG8025RR2 and AG8025conv, we took upregulated DEG genes ( $n=2460$ ) that were common between EdgeR and Deseq2 as well as down-regulated genes common between EdgeR/Deseq2, of these were 55.1% upregulated and 44.9% downregulated. Between the variety pair AG9045RR2 and AG9045conv,  $n=387$  genes were differentially expressed (79.6% upregulated and 20.4% downregulated). This number corresponded to  $\sim 6.21\%$  (for AG8025conv and AG8025RR2) and  $\sim 0.98\%$  (for AG9045conv and AG9045RR2) of all detected maize genes. Further investigation must be performed to convincingly explain why the variety pair AG8025RR2 has a significantly higher number of differentially expressed genes than the AG9045RR2 variety pair as it does not correlate to the total number of annotated genes in these libraries.

### Gene annotations and ontologies

In the analysis of RNA-seq experiments used to study several varieties, it is common to further analyze genes that are affected in both varieties. In our case, it turned out that the 27 upregulated genes and 26 downregulated genes were differentially expressed in the comparison of the AG8025RR2/AG8025conv and AG9045RR2/AG9045conv variety groups, respectively. To reduce artifacts, we decided to analyze not only two pairs of varieties but also performed several statistical evaluations. For this reason, in addition to individual gene evaluations, we determined matching annotations of GO ontologies and KEGG ontologies for the two GM/conventional groups. Because each ontology term contains clusters of multiple genes, the ontology results are less affected by single false-positive or false-negative results than DEGs. In the statistical evaluations of the GO ontologies, we used only those ontologies for further analysis that matched between AG8025RR2/AG8025conv and AG9045RR2/AG9045conv, which consisted of  $n=24$  ontologies with biological processes in both varieties. A REVIGO analysis enabled us to reduce this number to 16 GOs containing  $n=81$  genes (although other programs such as ClueGO are merging the groups differently). The result of REVIGO can be summarized as follows: the largest groups of ontologies were twelve different types of responses. Further ontologies contained three single-organism processes

and developmental processes involved in reproduction. When calculating KEGG annotations and even more importantly KEGG enrichments discovering significant pathways with  $p < 0.05$ , the results were different between the two variety pairs AG8025(RR2) and AG9045(RR2). Variety pair AG9045(RR2) contained less significant pathways ( $n = 2$ ) than variety AG8025(RR2) ( $n = 9$ ). The pathways with  $p < 0.05$  are shown in detail in Additional file 12 and Fig. 4 indicating some main components of the analysis as well as enrichment map of AG8025(RR2) and AG9045(RR2). Several aspects are worth to be emphasized. In variety pair AG9045(RR2) there was a cluster consisting of three pathways (zma00650: Butanoate metabolism; zma00280: Valine, leucine and isoleucine degradation; zma00410: beta-alanine metabolism). It is even more interesting that AG9045(RR2) had a second cluster containing the same pathway found also in AG8025(RR2) (zma01110: Biosynthesis of secondary metabolites).

In order to identify unintended effects arising from the insertion of the NK603 gene more in detail and to facilitate future investigations, we compared the results of SEA/REViGO and KEGG analysis and used the genes that occurred in both evaluations (see Table 5) for additional analyses. Thus, we have determined the similarity among the most important genes with a DAVID analysis and by means of Kappa values. In addition, we have listed the background genes of important KEGG pathways and the biochemical maps of these pathways in Additional files 14 and 15.

#### RT-qPCR

For a reliable interpretation of RNA-seq analyses, it is necessary to validate the results by additional molecular methods, for example by microarray or by RT-qPCR [3, 44, 61, 69]. La Paz et al. was able to show that a high number of DEGs analyzed with RNA-seq could be confirmed with both microarray and RT-qPCR [49]. We were able to confirm the results of two randomly selected differentially expressed genes using RT-qPCR as well. Evaluation of the analyses shows that the results are highly reliable. These two genes were randomly selected and are important for plant function. One gene with the ID GRMZM2G127948 codes for the protein caffeoyl-CoA O-methyltransferase 1, a key enzyme in lignin biosynthesis. Lignin provides mechanical strength to vascular tissues and protects plants from biotic stresses, including pathogen attack [51, 83]. The second gene with the ID AC204711.3\_FG003 is a gene (senescence-associated protein DIN) relevant for senescence-related processes. Senescence is involved in deterioration and is found in several parts of a plant or has functional characteristics

at the cellular level or may be involved in death rates or fecundity. The expression of these genes can be found in the articles of [16, 42, 47, 59, 66, 77, 78].

#### Interpretation of genetic diversity using PCA and heatmap

In the interpretation of our results, we are aware that there may be more genes that show differential expression in RNA-seq experiments of genetically modified organisms (GMOs) or when exposed to different environmental conditions. The results of the comparison of single GM crop varieties with corresponding near-isogenic varieties may only apply to the specific variety and to the conditions of the year when the variety was harvested. In order to assess variance structure of our data, we performed PCA and heatmap. The results confirm that we have robust data that differentiates the genotypes even by transgenic event although the highest variability is explained by genotype. AG8025RR2 and its near-isogenic variety have more DEGs than AG9045RR2 and its corresponding near-isogenic variety. Differences between the two near-isogenic varieties (AG9045 and AG8025) are higher than differences between AG9045RR2 and AG9045 but similar compared to the differences between AG8025RR2 and AG8025. This indicates that most of the transgenic effects are variety-dependent which is also implicated by the heatmap results. However, as the two pairs of varieties we examined have different genetic backgrounds, the DEGs occurring in both variety pairs could also indicate that some differences may be conserved among the two varieties. In order to assess the role of GM crops in more detail and to get a better insight into general effects of GM crops, such as unintended effects and pleiotropy [33, 50], further studies would have to be performed. For these investigations it would be important to carry out an analysis of additional NK603 varieties containing different genetic backgrounds and the experiments should be performed in more standardized and well defined environmental conditions [8]. To distinguish between environmental and genetic effects, the varieties would have to be cultivated at different times of the year, and the genetic relationship between GM varieties and comparators should be clear in more detail. However, it is difficult to implement this approach, as many varieties, and especially isogenic conventional lines, necessary for RNA-seq analyses of GM crops, are in the hands of corporations and have not been made available for research purposes in response to our enquiries.

### Interpretation of unintended genetic effects

Our results indicate that several unintended effects involved in different biological processes have occurred as a result of NK603 integration. Further single-gene analyses of the key genes and genes associated with affected biochemical pathways in different NK603 maize varieties and corresponding counterparts is necessary to interpret the role of the transgene and the biological significance of our findings.

Unintended genetic effects of GM plants have been described in other publications as well [2, 5–8, 12, 17, 19, 32, 35, 40, 45, 49, 58, 60, 84, 88]. According to these studies, transgenic inserts into the genome of one plant variety might affect the overall expression of other endogenous genes in the GM plant. Benevenuto et al. compared proteome profiles of herbicide-tolerant NK603 maize to near-isogenic non-GM maize under drought and herbicide stress and detected twenty differentially abundant proteins mainly assigned to energetic/carbohydrate metabolic processes. When comparing the NK603 maize and its non-GM near-isogenic variety under the same environmental conditions differences were identified in the levels of jasmonate, methyl jasmonate and cinnamic acid and in the abundance of 11 proteins [12]. This is similar to our results as we observed a high abundance of the “jasmonate ZIM domain-containing protein (ko13464)” across the DEGs which, together with their interacting partners, plays an essential role in orchestrating the cross talk between jasmonate and other hormone signaling pathways [89]. Agapito et al. analyzed the proteome profiles of stacked commercial maize hybrid containing insecticidal (BT = *Bacillus thuringiensis*) and herbicide tolerant traits (NK603) in comparison to the corresponding single event hybrids and non-GM conventional counterparts in the same genetic background as well as in comparison to a non-GM landrace variety under highly controlled growth conditions. Twenty-two proteins were differentially expressed in stacked and single GM events versus non-GM isogenic maize and a landrace variety. These proteins were mainly assigned to energy/carbohydrate and detoxification metabolism. Stacked GM genotypes were clustered together and distant from other genotypes analyzed by PCA. In addition, the varieties containing either BT or NK603 were clustered separately and clearly different from the non-GM varieties [2]. Arruda et al. and Herrera-Agudelo et al. detected significant differences in proteins, metalloproteins, enzymes and metals between transgenic soybeans harboring a RR insert and non-transgenic soybeans [5–7, 40].

However, many authors of these studies argue that the changes in gene expression, protein distribution and metabolite content caused by genetic modification are

less frequent than those caused by environmental factors or natural variations. In the study of Coll et al., natural variation explained most of the variability in gene expression among the samples, but the authors still emphasized that “transcriptional differences of conventionally bred varieties should be considered in the safety assessment of GM plants” [19]. When interpreting these studies, it should be borne in mind that genetic diversity in natural populations can be very large. Breeding or local populations generally show much lower genetic diversity than landraces and different types of wild relatives in maize [34]. Maize has been used as a crop for about 9,000 years [26] and its domestication resulted in a wide genetic diversity of native landraces [82]. During this period, considerable changes in the morphology and physiology of maize may have occurred. Thus, the quantity and quality of genetic effects and the ratio of genetic effects to effects of transgenic insertions may be very diverse in different populations. Specific aspects of risk assessment such as the selection of comparators have been discussed and developed by the EFSA GMO Panel [24].

### Secondary effects

Another important requirement for future RNA-seq analyses is to consider secondary effects. La Paz et al. described a small but significant delay in seed and plant maturation, which possibly influenced the functional annotation and expression of differentially expressed genes of MON810 plants [49]. To consider secondary effects in our analysis, we measured the weight of maize grains and compared them between NK603 and isogenic varieties. In these experiments, we did not find any consistent effects between the two varieties. However, we do not know whether there are other secondary effects between GM crops and conventional varieties. Thus, it would be important to carry out additional experiments, for which one could either evaluate certain parts of plants or otherwise analyze embryos as performed in the study of La Paz et al. [49].

### Conclusion

Overall, our data clearly demonstrate substantial differences between the analyzed transgenic varieties and their non-transgenic counterparts. PCA confirms a distinct difference between conventional and transgenic samples for variety AG8025 and a slight difference for the other variety pair. Several biological processes and metabolic pathways are modulated in the transgenic varieties. These differences indicate that several unintended effects have occurred as a result of NK603 integration. Heatmap data imply that most of the transgenic insert effects are variety-dependent. However, identified key genes involved in affected pathways of both variety pairs show

that transgenic independent effects cannot be excluded. Further research is necessary to clarify the role of internal and external influences on gene expression.

In general, our studies show that transcriptomic analysis is very useful to assess gene interactions, pleiotropic effects and unintended effects in transgenic crops. Thus, this technique may be a valuable tool for assessing genes that affect plant health or plant fitness, and serve as complementary safety analysis for the pre-market approval of GMOs. Even though RNA-seq has become the standard method for transcriptome analysis there are still analytical gaps that need to be taken into account, especially those related to the quantification of low levels transcripts [22]. However, with continuous developments of RNA-seq strategies, it is anticipated that more robust transcript identification will be able to be performed from longer reads. Thus, allowing a more accurate detection of individual, allele-specific biological variations and splice variants [61, 81].

In future investigations of NK603 varieties, it would be important to analyze genes of affected biochemical pathways more in detail to assess the influence of internal (such as variety or NK603 transgene) and external (such as environment) factors on the expression of these genes. Moreover, ontology clusters of this study should be evaluated more accurately and with other methods for example, different types of omics studies [48, 58]. Furthermore, analyzing stacked varieties [84] and the consideration of several transgenic events could be of great interest for the regulatory authorities.

## Supplementary information

**Supplementary information** accompanies this paper at <https://doi.org/10.1186/s12302-020-00412-8>.

**Additional file 1.** GenID, counts, and details of Deseq2 calculations for 8025 variety group. X75544, X75546, X75550 (AG8025RR2); X75553, X75556, X75559 (AG8025conv).

**Additional file 2.** GenID, counts, and details of EdgeR calculations for 8025 variety group.

**Additional file 3.** GenID, counts, and details of Deseq2 calculations for 9045 variety group. X30289, X30292, X30295 (AG9045conv); X30300, X30302, X30304 (AG9045RR2).

**Additional file 4.** GenID, counts, and details of EdgeR calculations for 9045 variety group.

**Additional file 5.** Deseq2 values for significant genes in 8025 variety group.

**Additional file 6.** EdgeR values for significant genes in 8025 variety group.

**Additional file 7.** Deseq2 values for significant genes in 9045 variety group.

**Additional file 8.** EdgeR values for significant genes in 9045 variety group.

**Additional file 9.** Result of GO ontologies for 8025 variety group.

**Additional file 10.** Result of GO ontologies for 9045 variety group.

**Additional file 11.** KEGG-KOBAS annotation.

**Additional file 12.** KEGG enrichment pathways.

**Additional file 13.** Results of weight measurements for maize grains.

**Additional file 14.** Background genes of enriched KEGG pathways of the key genes.

**Additional file 15.** Maps of the KEGG pathways.

## Abbreviations

DEGs: Significant differentially expressed genes; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; RT-qPCR: Reverse transcription quantitative PCR; GM(O): Genetically modified (organism); RR: Roundup Ready; EPSPS: 5-Enolpyruvylshikimate-3-phosphate synthase; HRM: High resolution melting; CIA: Chloroform-isoamylalcohol; RT: Reverse transcriptase; SEA: Singular enrichment analysis; RNA-seq: RNA-sequencing; FDR: False discovery rate; log<sub>2</sub> FC: Log<sub>2</sub>-fold change; BT: *Bacillus thuringiensis*.

## Acknowledgements

We would like to thank Marika Fodorova for the help with the manuscript. Moreover, we would like to express our gratitude to all reviewers for their helpful and detailed comments.

## Authors' contributions

CB, RH and AGH conceived and designed the experiments. CB supervised the project. DP, AD and SEB performed the experiments. CB, SA, DP and SEB analyzed the data. CB, SEB and SA wrote the paper. SA provided material for the experiments. All authors read and approved the final manuscript.

## Funding

The project was kindly supported by FEMtech (FFG) fellowships to Sina-Elisabeth Ben Ali, Agnes Draxler, and Diana Poelzl. RIAT-CZ project (ATCZ40) funded via Interreg V-A Austria—Czech Republic is gratefully acknowledged for the financial support of the measurements at the Vienna Biocenter Core Facilities, GmbH.

## Availability of data and materials

Sequencing data are available at NCBI:

Accession numbers for AG9045RR2 samples: SRR12052985, SRR12052988 and SRR12052989.

Accession numbers for AG9045 samples: SRR12052982, SRR12052983 and SRR12052984.

Accession numbers for AG8025RR2 samples: SRR12052978, SRR12052986 and SRR12052987.

Accession numbers for AG8025 samples: SRR12052979, SRR12052980 and SRR12052981.

## Ethics approval and consent to participate

Not applicable.

## Consent for publication

Not applicable.

## Competing interests

No conflicts of interest.

## Author details

<sup>1</sup> Institute for Food Safety, Austrian Agency for Health and Food Safety, Spargelfeldstrasse 191, 1220 Vienna, Austria. <sup>2</sup> Department of Nutritional Sciences, University of Vienna, Althanstraße 14, 1090 Vienna, Austria. <sup>3</sup> GenØk-Centre for Biosafety, Science Park, PB 6418, 9294 Tromsø, Norway.

Received: 6 April 2020 Accepted: 21 September 2020

Published online: 07 October 2020

## References

- Agapito-Tenfen SZ, Guerra MP, Wikmark O-G, Nodari RO (2013) Comparative proteomic analysis of genetically modified maize grown under different agroecosystems conditions in Brazil. *Proteom Sci* 11:46. <https://doi.org/10.1186/1477-5956-11-46>
- Agapito-Tenfen SZ, Vilperte V, Benevenuto RF, Rover CM, Traavik TI, Nodari RO (2014) Effect of stacking insecticidal cry and herbicide tolerance epsps transgenes on transgenic maize proteome. *BMC Plant Biol* 14:346–346. <https://doi.org/10.1186/s12870-014-0346-8>
- Agarwal P, Parida SK, Mahto A, Das S, Mathew IE, Malik N, Tyagi AK (2014) Expanding frontiers in plant transcriptomics in aid of functional genomics and molecular breeding. *Biotechnol J* 9:1480–1492. <https://doi.org/10.1002/biot.201400063>
- Anders S, Huber W (2010) Differential expression analysis for sequence count data. *Genome Biol* 11:R106–R106. <https://doi.org/10.1186/gb-2010-11-10-r106>
- Arruda M, Azevedo RA, Barbosa HS, Mataveli LRV, Oliveira SR, Arruda SCC, Gratão PL (2013) Comparative studies involving transgenic and non-transgenic soybean: what is going on? In: Board JE (ed) A comprehensive survey of international soybean research—genetics, physiology, agronomy and nitrogen relationships. Intech, Croatia, pp 583–613
- Arruda M, Galazzi R, Campos B, Herrera-Aguldelo MA, Arruda SCC, Azevedo R (2016) Soybean as a food source: comparative studies focusing on transgenic and nontransgenic soybean. In: Watson RR, Preedy VR (eds) *Genetically Modified Organisms in Food: Production, Safety, Regulation and Public Health*. Academic Press, Cambridge, pp 3–10
- Arruda SCC, Barbosa HS, Azevedo RA, Arruda MAZ (2013) Comparative studies focusing on transgenic through cp4EPSPS gene and non-transgenic soybean plants: an analysis of protein species and enzymes. *J Proteomics* 93:107–116. <https://doi.org/10.1016/j.jprot.2013.05.039>
- Barros E, Lezar S, Anttonen MJ, van Dijk JP, Rohlig RM, Kok EJ, Engel KH (2010) Comparison of two GM maize varieties with a near-isogenic non-GM variety using transcriptomics, proteomics and metabolomics. *Plant Biotechnol J* 8:436–451. <https://doi.org/10.1111/j.1467-7652.2009.00487.x>
- Behr C, Heck G, Hironaka C, You J (2012) Corn event PV-ZMGT32(nk603) and compositions and methods for detection thereof. United States Patent
- Ben Ali S-E et al (2018) Genetic and epigenetic characterization of the cry1Ab coding region and its 3' flanking genomic region in MON810 maize using next-generation sequencing. *Eur Food Res Technol* 244:1473–1485. <https://doi.org/10.1007/s00217-018-3062-z>
- Ben Ali SE, Madi ZE, Hochegger R, Quist D, Prewein B, Haslberger AG, Brandes C (2014) Mutation scanning in a single and a stacked genetically modified (GM) event by real-time PCR and high resolution melting (HRM) analysis. *Int J Mol Sci* 15:19898–19923. <https://doi.org/10.3390/ijms151119898>
- Benevenuto RF, Agapito-Tenfen SZ, Vilperte V, Wikmark O-G, van Rensburg PJ, Nodari RO (2017) Molecular responses of genetically modified maize to abiotic stresses as determined through proteomic and metabolomic analyses. *PLoS ONE* 12:e0173069. <https://doi.org/10.1371/journal.pone.0173069>
- Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Roy Stat Soc: Ser B* 57:289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
- Castan M, Ben Ali S-E, Hochegger R, Ruppitsch W, Haslberger AG, Brandes C (2017) Analysis of the genetic stability of event NK603 in stacked corn varieties using high-resolution melting (HRM) analysis and Sanger sequencing. *Eur Food Res Technol* 243:353–365. <https://doi.org/10.1007/s00217-016-2749-2>
- Chang S, Puryear J, Cairney J (1993) A simple and efficient method for isolating RNA from pine trees. *Plant Mol Biol Rep* 11:113–116. <https://doi.org/10.1007/BF02670468>
- Chen W et al (2016) Comparative and parallel genome-wide association studies for metabolic and agronomic traits in cereals. *Nat Commun* 7:12767. <https://doi.org/10.1038/ncomms12767>
- Cheng KC, Beaulieu J, Iquira E, Belzile FJ, Fortin MG, Strömvik MV (2008) Effect of transgenes on global gene expression in soybean is within the natural range of variation of conventional cultivars. *J Agric Food Chem* 56:3057–3067. <https://doi.org/10.1021/jf073505i>
- Cho K et al (2008) Integrated transcriptomics, proteomics, and metabolomics analyses to survey ozone responses in the leaves of rice seedling. *J Proteome Res* 7:2980–2998. <https://doi.org/10.1021/pr800128q>
- Coll A, Nadal A, Collado R, Capellades G, Kubista M, Messeguer J, Pla M (2010) Natural variation explains most transcriptomic changes among maize plants of MON810 and comparable non-GM varieties subjected to two N-fertilization farming practices. *Plant Mol Biol* 73:349–362. <https://doi.org/10.1007/s11103-010-9624-5>
- Coll A et al (2009) Gene expression profiles of MON810 and comparable non-GM maize varieties cultured in the field are more similar than are those of conventional lines. *Transgenic Res* 18:801–808. <https://doi.org/10.1007/s11248-009-9266-z>
- Coll A, Nadal A, Palau-delmas M, Messeguer J, Mele E, Puigdomenech P, Pla M (2008) Lack of repeatable differential expression patterns between MON810 and comparable commercial varieties of maize. *Plant Mol Biol* 68:105–117. <https://doi.org/10.1007/s11103-008-9355-z>
- Conesa A et al (2016) A survey of best practices for RNA-seq data analysis. *Genom Biol* 17:13. <https://doi.org/10.1186/s13059-016-0881-8>
- Costa-Silva J, Domingues D, Lopes FM (2017) RNA-Seq differential expression analysis: an extended review and a software tool. *PLoS ONE* 12:e0190152. <https://doi.org/10.1371/journal.pone.0190152>
- Devos Y et al (2014) EFSA's scientific activities and achievements on the risk assessment of genetically modified organisms (GMOs) during its first decade of existence: looking back and ahead. *Transgenic Res* 23:1–25. <https://doi.org/10.1007/s11248-013-9741-4>
- Dobin A et al (2012) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29:15–21. <https://doi.org/10.1093/bioinformatics/bts635>
- Doebley J (1990) Molecular evidence and the evolution of maize. *Econ Bot* 44:6–27
- Dündar F, Skrabanek L, Zumbo P (2015) Introduction to differential gene expression analysis using RNA-seq. Weill Cornell Medical College, New York
- Eckerstorfer M, Miklau M, Gaugitsch H (2014) Biosafety considerations for New Plant Breeding Techniques. Rep-0477. Environment Agency Austria, Vienna
- EFSA (2003) Opinion of the scientific panel on genetically modified organisms on a request from the commission related to the safety of foods and food ingredients derived from herbicide tolerant genetically modified maize NK603, for which a request for placing on the market was submitted under Article 4 of the Novel Food Regulation (EC) No 258/97 by Monsanto (QUESTION NO EFSA-Q-2003-002). *EFSA J* 9:1–14
- EFSA (2010) Guidance on the environmental risk assessment of genetically modified plants. *EFSA J*. <https://doi.org/10.2903/j.efsa.2010.1879>
- EFSA (2018) EFSA Scientific Colloquium 24—omics in risk assessment: state of the art and next steps. In: *EFSA Scientific Colloquium 24—omics in risk assessment: state of the art and next steps*. EFSA supporting publication, Berlin, pp 1–30. doi:<https://doi.org/10.2903/sp.efsa.2018.EN-1512>
- Ei Ouakfaoui S, Miki B (2005) The stability of the Arabidopsis transcriptome in transgenic plants expressing the marker genes nptII and uidA. *Plant J* 41:791–800. <https://doi.org/10.1111/j.1365-313X.2005.02350.x>
- Fagan J, Antoniou M, Robinson C (2014) *GMO Myths and Truths*, 2nd edn. Earth Open Source, London
- Gore MA et al (2009) A first-generation haplotype map of maize. *Science* 326:1115–1117. <https://doi.org/10.1126/science.1177837>
- Gregersen PL, Brinch-Pedersen H, Holm PB (2005) A microarray-based comparative analysis of gene expression profiles during grain development in transgenic and wild type wheat. *Transgenic Res* 14:887–905. <https://doi.org/10.1007/s11248-005-1526-y>
- Gulli M, Salvatori E, Fusaro L, Pellacani C, Manes F, Marmiroli N (2015) Comparison of drought stress response and gene expression between a GM maize variety and a near-isogenic non-GM variety. *PLoS ONE* 10:e0117073. <https://doi.org/10.1371/journal.pone.0117073>
- Harrigan GG et al (2016) Evaluation of metabolomics profiles of grain from maize hybrids derived from near-isogenic GM positive and negative segregant inbreds demonstrates that observed differences cannot be attributed unequivocally to the GM trait. *Metabolomics* 12:82–82. <https://doi.org/10.1007/s11306-016-1017-6>
- Haslberger AG (2003) Codex guidelines for GM foods include the analysis of unintended effects. *Nat Biotechnol* 21:739–741. <https://doi.org/10.1038/nbt0703-739>

39. Heck GR et al (2005) Development and characterization of a CP4 EPSPS-based, glyphosate-tolerant corn event. *Crop Sci* 45:329–339. <https://doi.org/10.2135/cropsci2005.0329>
40. Herrera-Agudelo MA, Miró M, Arruda MAZ (2017) In vitro oral bioaccessibility and total content of Cu, Fe, Mn and Zn from transgenic (through cp4 EPSPS gene) and nontransgenic precursor/successor soybean seeds. *Food Chem* 225:125–131. <https://doi.org/10.1016/j.foodchem.2017.01.017>
41. Herrero M, Ibanez E, Martin-Alvarez PJ, Cifuentes A (2007) Analysis of chiral amino acids in conventional and transgenic maize. *Anal Chem* 79:5071–5077. <https://doi.org/10.1021/ac070454f>
42. Hill-Skinner S (2018) Genetic and environmental control of lignin biosynthesis and C emission from crop stover. Iowa State University, Ames
43. ISAAA (2018) Global Status of Commercialized Biotech/GM Crops in 2018: Biotech Crops Continue to Help Meet the Challenges of Increased Population and Climate Change. ISAAA. <https://www.isaaa.org/purchasepublications/itemdescription.asp?itemType=BRIEFS&Control=IB054-2018>. Accessed 01 Jan 2020
44. Jorrin-Novo JV (2014) Plant Proteomics methods and protocols. In: Jorrin-Novo JV, Komatsu S, Weckwerth W, Wienkoop S (eds) *Plant proteomics*. Springer, New York, pp 3–13
45. Kogel K-H et al (2010) Transcriptome and metabolome profiling of field-grown transgenic barley lack induced differences but show cultivar-specific variances. *Proc Natl Acad Sci USA* 107:6198–6203. <https://doi.org/10.1073/pnas.1001945107>
46. Kohli A, Twyman RM, Abranches R, Wegel E, Stoger E, Christou P (2003) Transgene integration, organization and interaction in plants. *Plant Mol Biol* 52:247–258. <https://doi.org/10.1023/a:1023941407376>
47. Kost M (2014) Maize and Sunflower of North America: Conservation and Utilization of Genetic Diversity. Dissertation, Ohio State University
48. Kuo T-C, Tian T-F, Tseng YJ (2013) 3Omics: a web-based systems biology tool for analysis, integration and visualization of human transcriptomic, proteomic and metabolomic data. *BMC Syst Biol* 7:64–64. <https://doi.org/10.1186/1752-0509-7-64>
49. La Paz JL, Pla M, Centeno E, Vicient CM, Puigdomènech P (2014) The use of massive sequencing to detect differences between immature embryos of MON810 and a comparable non-GM maize variety. *PLoS ONE* 9:e100895–e100895. <https://doi.org/10.1371/journal.pone.0100895>
50. Ladics GS et al (2015) Genetic basis and detection of unintended effects in genetically modified crop plants. *Transgenic Res* 24:587–603. <https://doi.org/10.1007/s11248-015-9867-7>
51. Liu Q, Luo L, Zheng L (2018) Lignins: biosynthesis and biological functions in plants. *Int J Mol Sci* 19:335. <https://doi.org/10.3390/ijms19020335>
52. Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the  $2^{-\Delta\Delta C(T)}$  method. *Methods* 25:402–408. <https://doi.org/10.1006/meth.2001.1262>
53. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15:550. <https://doi.org/10.1186/s13059-014-0550-8>
54. Lowe R, Shirley N, Bleackley M, Dolan S, Shafee T (2017) Transcriptomics technologies. *PLoS Comput Biol* 13:e1005457–e1005457. <https://doi.org/10.1371/journal.pcbi.1005457>
55. Manoli A, Sturaro A, Trevisan S, Quaggiotti S, Nonis A (2012) Evaluation of candidate reference genes for qPCR in maize. *J Plant Physiol* 169:807–815. <https://doi.org/10.1016/j.jplph.2012.01.019>
56. Mao X, Cai T, Olyarchuk JG, Wei L (2005) Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. *Bioinformatics* 21:3787–3793. <https://doi.org/10.1093/bioinformatics/bti430>
57. McHugh ML (2012) Interrater reliability: the kappa statistic. *Biochem Med (Zagreb)* 22:276–282
58. Mesnage R et al (2016) An integrated multi-omics analysis of the NK603 Roundup-tolerant GM maize reveals metabolism disturbances caused by the transformation process. *Sci Rep* 6:37855. <https://doi.org/10.1038/srep37855>
59. Meyer A (2015) A bioinformatic analysis of genetic factors affecting primary root growth in *Zea mays*. Dissertation, The University of Guelph, Guelph
60. Montero M, Coll A, Nadal A, Messeguer J, Pla M (2011) Only half the transcriptomic differences between resistant genetically modified and conventional rice are associated with the transgene. *Plant Biotechnol J* 9:693–702. <https://doi.org/10.1111/j.1467-7652.2010.00572.x>
61. Mutz K-O, Heilkenbrinker A, Lönne M, Walter J-G, Stahl F (2013) Transcriptome analysis using next-generation sequencing. *Curr Opin Biotechnol* 24:22–30. <https://doi.org/10.1016/j.copbio.2012.09.004>
62. Neumann G, Brandes C, Joachimsthaler A, Hoegger R (2011) Assessment of the genetic stability of GMOs with a detailed examination of MON810 using Scorpion probes. *Eur Food Res Technol* 233
63. OECD (2015) Molecular characterisation of plants derived from modern biotechnology, vol 2. OECD, Paris
64. Pfaffl M (2004) A-Z of quantitative PCR. International University Line, La Jolla
65. Poerschmann J, Gathmann A, Augustin J, Langer U, Górecki T (2005) Molecular composition of leaves and stems of genetically modified Bt and near-isogenic non-Bt maize—characterization of lignin patterns. *J Environ Qual* 34:1508–1518. <https://doi.org/10.2134/jeq2005.0070>
66. Poloni A (2015) Investigation of host specificity mechanisms of *Sporisorium reilianum* in maize and sorghum. Aachen University, Aachen
67. Rapaport F et al (2013) Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data. *Genome Biol* 14:R95–R95. <https://doi.org/10.1186/gb-2013-14-9-r95>
68. Rasmussen R (2001) Quantification on the LightCycler. *Rapid Cycle Real-Time PCR*. Springer, Berlin, Heidelberg
69. Rey M-D et al (2019) Recent advances in MS-based plant proteomics: proteomics data validation through integration with other classic and -omics approaches. In: Cánovas F, Lüttge U, Leuschner C, Riusueño MC (eds) *Progress in botany*, vol 81. Springer, Cham
70. Ridley WP, Sidhu RS, Pyla PD, Nemeth MA, Breeze ML, Astwood JD (2002) Comparison of the nutritional profile of glyphosate-tolerant corn event NK603 with that of conventional corn (*Zea mays* L.). *J Agric Food Chem* 50:7235–7243. <https://doi.org/10.1021/jf0205662>
71. Rischer H, Oksman-Caldentey KM (2006) Unintended effects in genetically modified crops: revealed by metabolomics? *Trends Biotechnol* 24:102–104. <https://doi.org/10.1016/j.tibtech.2006.01.009>
72. Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26:139–140. <https://doi.org/10.1093/bioinformatics/btp616>
73. Saxena D, Stotzky G (2001) Bt corn has a higher lignin content than non-Bt corn. *Am J Bot* 88:1704–1706
74. Schnell J et al (2015) A comparative analysis of insertional effects in genetically engineered plants: considerations for pre-market assessments. *Transgenic Res* 24:1–17. <https://doi.org/10.1007/s11248-014-9843-7>
75. Schurch NJ et al (2016) How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use? *RNA* 22:839–851. <https://doi.org/10.1261/rna.053959.115>
76. Seyednasrollah F, Laiho A, Elo LL (2015) Comparison of software packages for detecting differential expression in RNA-seq studies. *Brief Bioinform* 16:59–70. <https://doi.org/10.1093/bib/bbt086>
77. Shimada Y, Wu G-J, Watanabe A (1998) A protein encoded by din1, a dark-inducible and senescence-associated gene of radish, can be imported by isolated chloroplasts and has sequence similarity to sulfide dehydrogenase and other small stress proteins. *Plant Cell Physiol* 39:139–143. <https://doi.org/10.1093/oxfordjournals.pcp.a029350>
78. Song Y et al (2016) Association of the molecular regulation of ear leaf senescence/stress response and photosynthesis/metabolism with heterosis at the reproductive stage in maize. *Sci Rep* 6:29843. <https://doi.org/10.1038/srep29843>
79. Supek F, Bosnjak M, Skunca N, Smuc T (2011) REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS ONE* 6:e21800. <https://doi.org/10.1371/journal.pone.0021800>
80. Tian T, Liu Y, Yan H, You Q, Yi X, Du Z, Xu W, Su Z (2017) agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update. *Nucleic Acids Res* 45(W1):W122–W129. <https://doi.org/10.1093/nar/gkx382>
81. Tilgner H, Grubert F, Sharon D, Snyder MP (2014) Defining a personal, allele-specific, and single-molecule long-read transcriptome. *Proc Natl Acad Sci* 111:9869–9874. <https://doi.org/10.1073/pnas.1400447111>

82. Vielle-Calzada J-P et al (2009) The palomero genome suggests metal effects on domestication. *Science* 326:1078–1078. <https://doi.org/10.1126/science.1178437>
83. Wang GF, Balint-Kurti PJ (2016) Maize homologs of CCoAOMT and HCT, two key enzymes in lignin biosynthesis, form complexes with the NLR Rp1 protein to modulate the defense response. *Plant Physiol* 171:2166–2177. <https://doi.org/10.1104/pp.16.00224>
84. Wang XJ, Zhang X, Yang JT, Wang ZX (2018) Effect on transcriptome and metabolome of stacked transgenic maize containing insecticidal cry and glyphosate tolerance epsps genes. *Plant J* 93:1007–1016. <https://doi.org/10.1111/tpj.13825>
85. Xie C, Mao X, Huang J, Ding Y, Wu J, Dong S, Kong L, Gao G, Li CY, Wei L (2011) KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Res* 39:W316–W322. <https://doi.org/10.1093/nar/gkr483>
86. Yu G, Wang L, Han Y, He Q (2012) clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 16(5):284–287. <https://doi.org/10.1089/omi.2011.0118>
87. Yu G (2020) enrichplot: visualization of functional enrichment result. R package version 1.8.1, <https://github.com/GuangchuangYu/enrichplot>.
88. Zanatta CB, Benevenuto RF, Nodari RO, Agapito-Tenfen SZ (2020) Stacked genetically modified soybean harboring herbicide resistance and insecticide rCry1Ac shows strong defense and redox homeostasis disturbance after glyphosate-based herbicide application. *Environ Sci Eur*. <https://doi.org/10.1186/s12302-020-00379-6>
89. Zhou X et al (2015) A maize Jasmonate Zim-domain protein, ZmJAZ14, associates with the JA, ABA, and GA signaling pathways in transgenic arabidopsis. *PLoS ONE* 10:e0121824. <https://doi.org/10.1371/journal.pone.0121824>
90. Zolla L, Rinalducci S, Antonioli P, Righetti PG (2008) Proteomics as a complementary tool for identifying unintended side effects occurring in transgenic maize seeds as a result of genetic modifications. *J Proteome Res* 7:1850–1861. <https://doi.org/10.1021/pr0705082>

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)

---