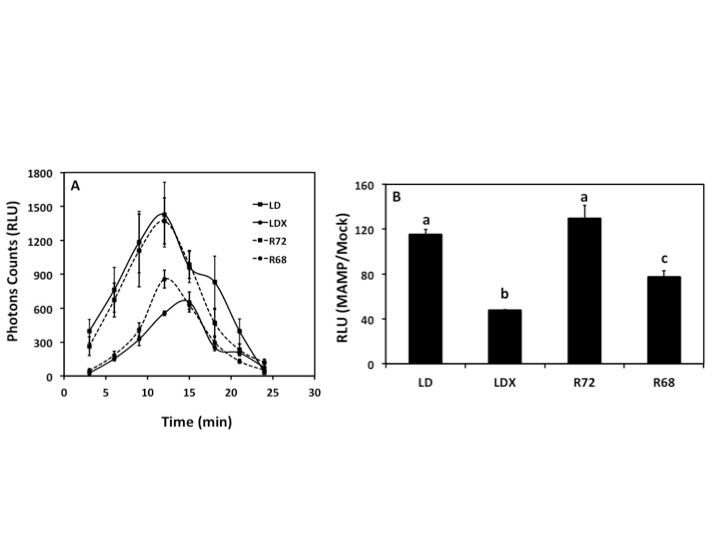
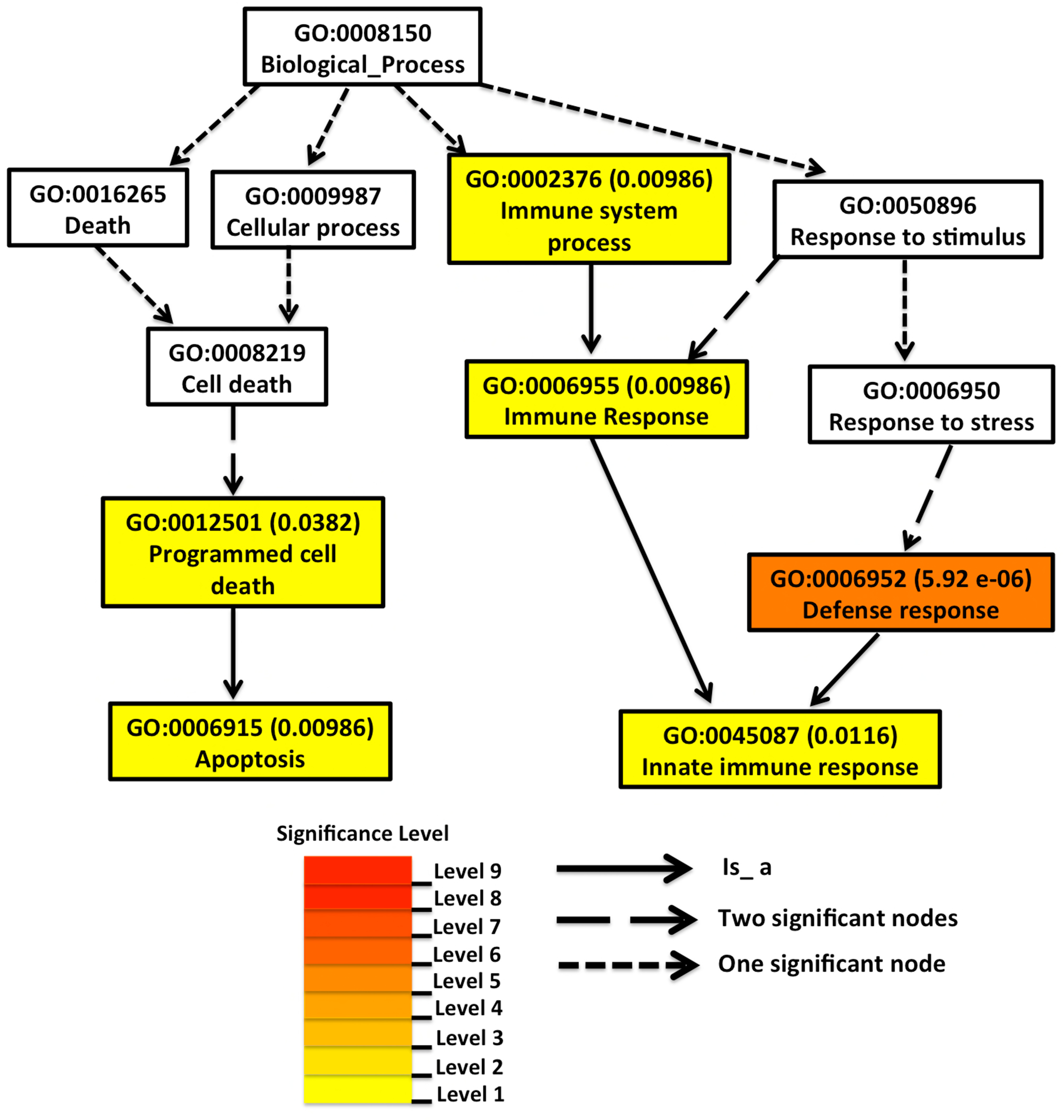
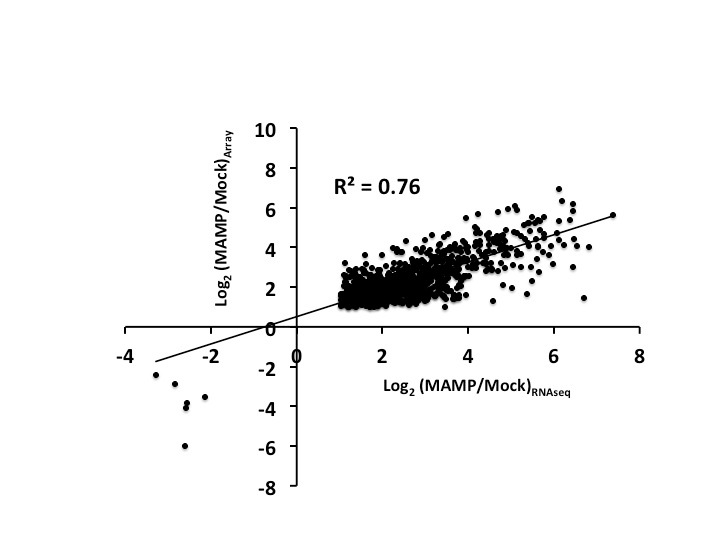
**Supplementary Material**

****

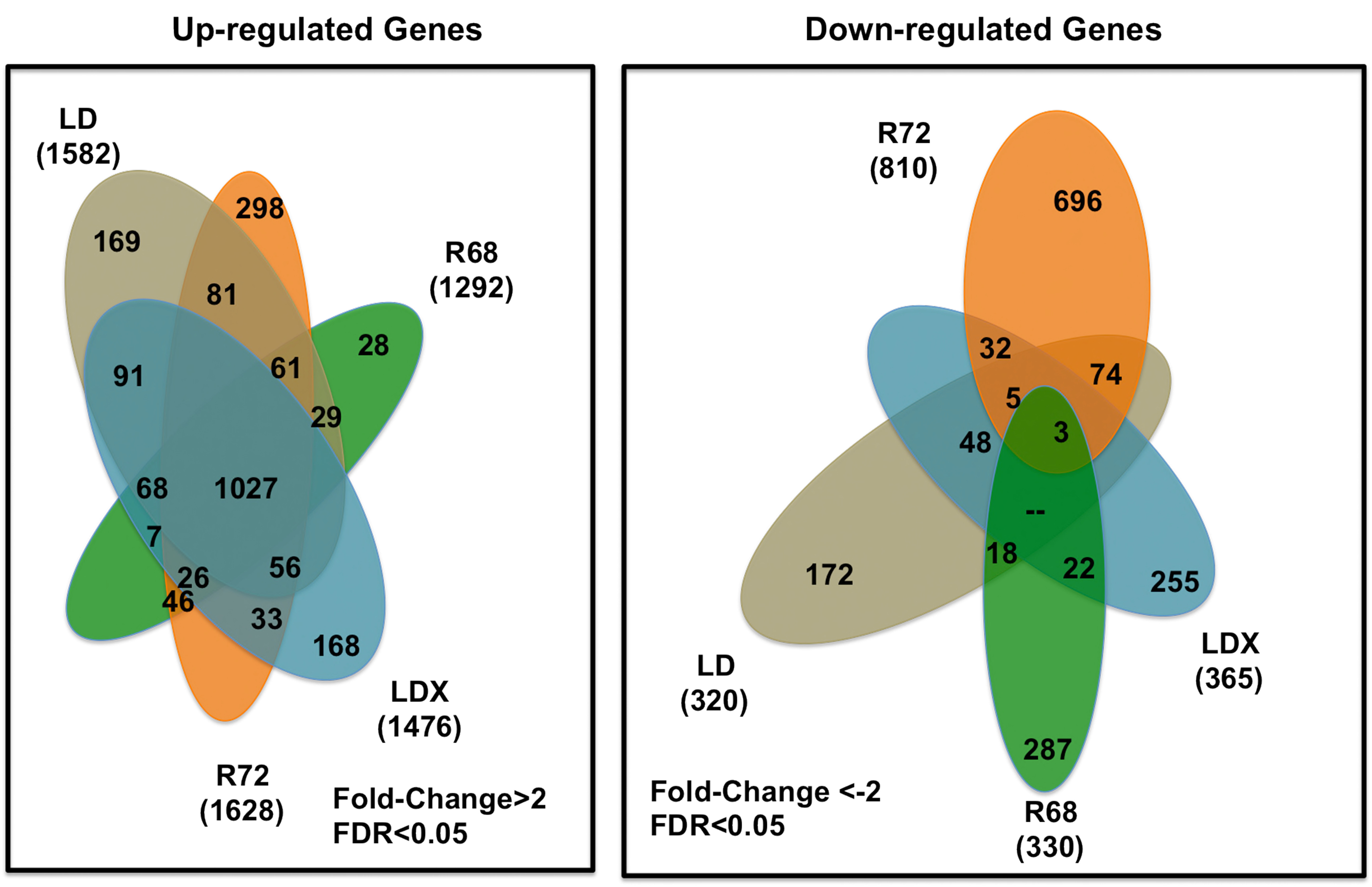
**Fig. S1:** Oxidative burst triggered by a mixture of 1 µM flg22 + 50 µg/ml chitin in the parental lines LD, LDX, as well as in the F4 lines R72, and R68 leaf discs measured in relative luminescence units (RLU). (A) Time course of the MAMP-triggered oxidative burst. Each data point represents the average of three biological replicates with three technical replicates. Error bars represent ± SE of the average. (B) Total ROS produced during the oxidative burst over 30 min of MAMP treatment. Results are the average ± standard error (s.e.) of the ratio between MAMP/Mock from three biological replicates with three technical replicates for each. Within each diagram, values sharing the common letters are not significantly different at *P* ≤ 0.05 by ANOVA.

****

**Fig. S2**:Enriched functional terms of the differentially regulated genes identified in the pairwise comparison between the parent LD and the F4 line R72. The GO terms statistically enriched are highlighted in yellow or orange. The GO term number and the P-values of the significantly enriched processes are indicated inside in each square.



**Fig S3.** Differential expression in the parental lines LD and LDX is concordant between mRNA-seq and Gene Chip Array. Estimates of differentially expression obtained using RNA-seq and Soybean Gene Chip Array (Array) are correlated. Point represents genes, located based on estimates of differential expression using RNA-seq and Soybean Gene Chip Array. The line corresponds to perfect agreement between both technologies.

****

**Fig. S4**: Flower diagram showing number of overlapping and non-overlapping MAMP-responsive genes among LD, LDX, R72, and R68. Differentially expressed genes in each genotype were identified by a generalized Poisson linear mixed-effects model at FDR < 0.05, with additional cutoff of 2-fold ratio in pairwise comparisons (MAMP-treated leaves *vs* Mock-treated leaves). Over- and non-overlapping genes were identified after a pairwise comparison between genotypes (i.e., LD vs LDX). Genes belong to each intersection are provided in Table S4.

**Material and Methods**

**Plant Material and MAMP treatments**

Seeds were surface sterilized [30% bleach, rinsed several times with sterile, double-distillated H2O (ddiH2O)], planted in soil-less medium (SogeMix “SM-2”-general purpose growing medium), and germinated in growth chambers (190 mol PAR m-2s-1, 230C, 60% R.H., and 16 h photoperiod). Trifoliate leaves from three-week-old plants were detached and then vacuumed infiltrated with ddiH2O for 2 min. Water-infiltrated trifoliate leaves from five different plants of each genotype were pooled and cut into approximately 1 cm2 slices. An equal amount of leaf slices (~30 slices) from each treatment were transferred into two different Petri dishes and then floated overnight on autoclaved ddiH2O. Water was removed from both Petri dishes and replaced with 5 ml of MAMP solution [1 µM flagellin 22 (flg22; Genescript) and 50 µg of crab shell chitin (referred to as chitin in this study; Sigma-Aldrich)], or 5 ml of mock solution [autoclaved ddiH2O plus equivalent amount of dimethyl sulfoxide (DMSO; Fisher Scientific). DMSO was included since it was contained in the solution used to dissolve the flg22 peptide]. After a 30 min treatment, mock- and MAMP-treated leaf slices were harvested into different tubes and immediately frozen in liquid nitrogen. Samples were stored at -80oC until use. All procedures described above were performed under dark conditions.

**Oxidative burst measurement using soybean leaf discs**

ROS production by leaf tissue was assayed by H2O2-dependent luminescence of luminol (Keppler et al., 1989). Under dark conditions, four 4- mm leaf discs per treatment from soybean water-infiltrated trifoliolate leaves were floated for 16 h in 200 L of autoclaved double-distilled water in a 48-well plate. Continuing under dark conditions, water was removed and replaced with 150 L of MAMP reaction buffer (20 mM luminol [Sigma-Aldrich], 1 mg horse- radish peroxidase [Sigma-Aldrich], 1 mM flg22 [Genescript] and 50 mg chitin [Sigma- Aldrich]) or mock reaction buffer (20 mM luminol, 1 g horseradish peroxidase, and 2 L of DMSO [Fisher Scientific]). Luminescence was immediately recorded over 30 min using a CCD photon- counting camera (Photek 216). The luminescence intensity of individual wells was calculated using Photek IFS32 software (Photek 216). The luminescence intensity data from MAMP-treated leaf discs were normalized using intensity values from mock-treated plants. Parental lines and F4 lines had three biological replicates with three technical replicates in each.

**Preparation of mRNA-seq libraries.**

Total RNA was isolated from 0.5 g of MAMP (see above)- or Mock-treated leaves using Trizol and subsequently purified using chloroform precipitation. Non-strand-specific mRNA-seq libraries were generated from 4 µg of total RNA and prepared by using the TruSeq RNA sample Prep Kit (Illumina) according to the manufacturer’s instructions.

**Mapping and Processing of mRNA-seq reads**

Initial base calling and quality filtering of the mRNA-Seq reads generated with the Illumina analysis pipeline (in the FASTQ format) were performed using a custom Perl script and the default parameters of the Illumina pipeline (http://www.illunina.com). Additional filtering for bad-quality bases (any base with Phred quality score values larger than 20 percentile was considered as a base with good quality) in each individual, and read size (>40 bp) was performed using the FASTA/Q Trimmer command of the FASTX-toolkit available in the FastQC software package (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). mRNA-seq reads with good quality were aligned to the *Glycine max* reference genome (Gmax1.01 release) using Tophat (version 1.4.1) (Trapnell et al., 2009). The genome indexes for Tophat were built using bowtie-build command of bowtie (version 0.12.7) (Langmead et al., 2009) using the reference genome file as input. Tophat was then run with default parameters to map the trimmed- and filtered-reads for each library to the reference genome. Tophat was supplied with the reference GTF file using the –G option and replicates of each condition/sample were mapped independently to improve alignment sensitivity and accuracy for further analysis. For analysis of protein-coding genes, only uniquely mapping reads were used.

**Identification of differentially expressed genes**

Low-count reads with a total sum fewer than 10 were filtered out [26] before a Poisson linear mixed-effects model (Blekhman et al., 2009) was applied to the raw read counts separately for each gene using the software R/lme4 package (2.10.0 version; http://lme4.r-forge.r-project.org/) with the library size as the offset value to make the comparison across different samples comparable. Each generalized Poisson linear mixed model includes the genotype effect, treatment effect, the interaction between the genotype and treatment effects, and the random biological replicate effect, as well as random pot effect accounting for the correlation between observations that share the same pot. Using the fitted values from the gene-specific model (also called the conditional mean of the observations; described above) as expression values, a likelihood ratio tests were conducted to identify differentially expressed genes between the treatment and control groups for each of the four genotypes, as well as between genotypes either control or treatment group (i.e., LDcontrol *vs* LDXcontrol). P-values for the likelihood ratio tests were obtained, and an adjusted-P value (Storey and Tibshirani, 2003) was then computed to produce lists of differentially expressed genes with an estimated FDR of 5%. Among these significantly differentially expressed genes, genes with a fold change above 2 were further considered.

**References**

Blekhman R, Marioni JC, Zumbo P, Stephens M, Gilad Y (2010) Sex-specific and lineage-specific alternative splicing in primates. *Genome Research* 20: 180-189.

Kepler LD, Baker CJ, Atkinson MM (1989) Active oxygen production during a bacteria-induced hypersensitive reaction in tobacco suspension cells. *Phytopathology* 79: 974-978.

Langmead B, Trapnell C, Pop M (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* 10: R25.

Storey JD, Tibshirani R (2003) Statistical significance for genomewide studies. *Proceeding of the National Academy of Sciences of the United States of America* 100: 9440-9445.

Trapnell C, Pachter L, Sazlberg SL (2009) TopHat: discovering splice junctions with RNA-seq. *Bioinformatics* 25: 1105-1111.