Supplementary Material

Supplementary Table S1: Absolute values for the Affymetrix array analyses for extended night samples. Published datasets of diurnal changes in Colo0 and *pgm* (Bläsing et al., 2005) are also included. The entire dataset was normalized using RMA. Data from the diurnal cycle of the WT is given indicated by a "d" and the number of hours into the diurnal cycle, where zero hours represents the end of the night. Data from *pgm* was labeled accordingly but starts with a "p". Extended night datasets are labeled by "x" and the number of hours in the extended night. An additional point xn_2 is data from a plant harvested from the same series but two hours before the end of the night.

Supplementary Table S2: LIMMA Analysis of the array data. For the experiments of the diurnal cycles of WT and *pgm* as well as for experiments form the extended night the RMA expression estimates were calculated using BioConductor. Subsequently, a linear model was fitted fitted to the extended night dataset using the limma package. Here, a significant difference to the end of the night for each data point was tested after adjusting down the p-values of genes, using the Benjamini Hochberg procedure (1995) and then across contrasts using the "hierarchical" method from the limma package. Each time point is labeled by an "x" for extended night and the hours into the extended nigh as well as by a "-x0" to show that all values are contrasted to the end of the night.

Supplementary Table S3: Weightings of genes in the principle components of a PCA of the combined dataset including the extended night, diurnal cycle in wild-type Col0 and *pgm*, and 4 h illumination with ambient or low [CO_2] (see Fig. 3 for the PCA plot). The data for the extended night is from the present paper, and for the other treatments from Bläsing et al. (2005).

Supplementary Table S4: Lists, groupings and responses of genes in the test sets for clock, light- and C-responsive genes.

This Supplemental Table provides background information to the gene clustering in Figs. 4-5, Table 1, Table III and Fig 7, and the plots in Figs. 8-11. It provides lists of the genes in each test set, indicators to allow them to be sorted, and documents the responses of their transcript levels in the experimental dataset that were used to identify them as members of the test set, and their

responses of their transcript levels during a diurnal cycle and an extended night in Col0 and a diurnal cycle in *pgm*

The test sets of genes were defined as follows:

Clock: A test set of 604 genes for the clock response was identified from a dataset was downloaded for the NASCArray database (Craigon et al. 2004) that describes the changes of gene expression during the second and third days after transfer of seedlings growing on 3% sucrose from a 12h light/ 12h dark cycle into continuous light. Samples were collected 2, 6, 10, 14, 18 and 22 h into the subjective day. The 1000 genes with the highest variance were subjected to a Fisher's g-test to test for periodicity, as implemented in the Genets package (Wichert et al., 2004). The 635 genes that were called significant ($p<0.05$) after a Benjamini-Hochberg (1995) correction were retained.

Carbon: The test sets of 484 C-induced genes and 383 C-repressed genes were identified as described in the main text and the legend of supplemental Table S4.

Light: The test sets of ca. 200 light-induced and 200 light-repressed genes correspond to the most strongly responding genes in a dataset from AtGenExpress. Seedlings were grown in the dark for 7 days and then illuminated for 4 hours with white light.

The Supplemental Table lists all these genes with their Affymetrix identifiers, AGI codes and TAIR6 annotations as obtained from MapMan. It provides

- numbered columns that identify genes according to the set(s) they belong to. Genes in the circadian clock set are identified in the corresponding colum with '1'. For light (photomorph) and C, the corresponding columns contain the symbols 1 and -1 to distinguish genes that are induced and repressed. A subset of the C-induced and C-repressed genes identified in a separate column as rapid responders. Transcripts for these genes changed by >2-fold within 30 min after adding sucrose to C-starved seedlings (Osuna et al., 2007).
- information about the time at which the transcript peaks in a free-running cycle.
- information about the cluster the gene was sorted into by Kmeans-clustering. Plots of the responses of these classes are given for the all of the C-responsive sets, the rapidly responding C-regulated genes and the light-responsive genes are given in Figs. 4, 5 and 7, respectively

The Supplemental Table also provides for each gene:

1. Datasets for treatments that allow the genes assignments to be verified:

Clock responses: The data downloaded from the NASCArray database described above. The data are given as absolute signals, and as ratios normalised on the estimated value at the end of the 24 h cycle. The samples were collected 2, 6, 10, 14, 18 and 22 h into the subjective day. The value at 48h (end of the second subjective day in the free-running cycle) 24 h was calculated from the signals at 46 and 50 h, assuming a linear change between these two time points.

C-responses. Two treatments are given:
- Endogenous changes of C in 5-week-old rosettes: datasets for 4 h illumination at the end of the night in 50 ppm [CO_2] and with 350 ppm [CO_2] (Bläsing et al., 2005). The data are given as absolute signals, and after normalisation on a control that was harvested after 4h extension of the night
- C-starvation and re-supply in seedlings. Seedlings growing in liquid culture in continuous low light were starved of exogenous C for 2 days were starved of C, and then resupplied with 15 mM sucrose for 30 min and 3 hours. Data for seedlings in full nutrition (FN) with 0.5% sucrose is also given. The results are given as absolute signals, and normalized on the C-starved seedlings.

Light responses. Two treatment are given
- Response to light in 5week-old light-grown rosettes. Plants were transferred to 50 ppm [CO_2] iat the end of the might and left in the dark for 4h or illuminated for 4h at 50 ppm [CO_2] (Bläsing et al., 2005). The data are given as absolute signals, and normalised on the signal in wild-type Col0 at the end of the night.
- Response of etiolated seedling to light. Seedlings were grown in the dark for 7 days and then illuminated for 4 hours with white light. The data were downloaded from AtGenExpress. They are given as absolute signals, and normalised on the value in etiolated seedlings.

2: Datasets that display the response in complex physiological sequences

Diurnal cycle in wild-type Col0. The data are from Bläsing et al. (2005) and represent the mean of independent biological triplicates sampled 4, 8 and 12 h into the light period, and 4, 8 and 12h (the latter is defined as ‘End of Night’ = EN) into the dark period.

Extended night in Col0 wild-type. The data are from the present article, and represent samples harvested at the end of the night, and 2, 4, 6, 8, 24 and 48 h into an extended night. The latter two time points are shown separately to emphasise the break in the time scale.

Diurnal cycle in pgm. The data are from Bläsing et al. (2005) and represent the mean of biological replicates at the start and end of the day, and single samples at other time points.

All three datasets are given as absolute values, and normalised on the values in wild-type Col0 at the end of the night.

Supplemental Table S5: Dataset to load into PageMan

Log2 fold changes for all time points of the diurnal series of the WT and the *pgm* mutant as well as for the extended night dataset, normed on the respective end of the night datpoint. These can be directly loaded into PageMan for analysing the changes based on functional MapMan classes.

Supplemental Table S6: Model errors

This table provides for each modelled transcripts, a qualitative and quantitative measure of overall performance in the form of correlation between true and modelled vales and Euclidean distance of the model from the truth. Also it provides variance of input and output as well as weighted inputs, to be able to filter genes out that are hardly reacting in the in put datasets. Finally functional MapMan classes are identified for all transcripts.

Supplementary Figure S1: Analysis of the extended night dataset. (A) Pair-wise scatter plots of the samples at the end of the night from the three replicate experiments. The slope and R^2 values are given on the plots. (B). Number of genes showing a statistically significant (FDR<0.05) increase of decrease of their transcript level at different times in an extended night treatment, compared to the end of the normal night. The datasets shown in Fig. 2A were subjected to a LIMMA analysis (see suppl. Table S2), and adjusted using the Benjamini-Hochberg correction. (C) Principal components analysis. Datasets were combined for three independent experiments, including one with samples collected at 0, 4, 6 and 8 h, one with samples collected after 0, 4, 8, 24 and 48 h, and one with samples collected after 0, 24 and 48 h (0 = end of the night. The integer indicates the time in hours into the extended night treatment.

Supplementary Figure S2: Clustering of samples from three extended night experiments, triplicated Col0 diurnal cycle treatments, *pgm* diurnal cycle and duplicated [CO2] treatments. . This plot analyses the same set of samples as is used for the PCA shown in Fig. 3. The samples were clustered by hierarchical cluster analysis using complete linkage. For details see the Methods section and the legend to Fig. 3.

Supplementary Figure S3: A comparison of gene weightings in the first principal component and the response of gene expression to sugars. (A) Response 3 hours after adding 15 mM sucrose to seedlings that had been starved for 2 days (data from Osuna et al., 2007). (B) Response 3 hours after adding 100 mM glucose to seedlings that had been starved for 2 days (data from Bläsing et al., 2005). The changes of expression are shown relative to the level in starved seedlings, and expressed on a log2 scale. (C) Frequency plot of the correlation coefficients calculated from 1000 shuffled datasets. (D) Response 30 min after adding 15mM sucrose to C-starved seedlings (Osuna et al., 2007). (E) Comparison of seedlings in full nutrient media with C-starved seedlings (Osuna et al., 2007). (F) Comparison of 5-week old Arabidopsis rosettes that were illuminated for 4 h starting form the end of the night with 5 350 compared 50 ppm [CO_2] (Bläsing et al., 2005).

Supplemental Figure S4. This provides additional documentation to Fig 6. (A) Functional categories related to photosynthesis, chloroplast biogenesis, pigment synthesis and nitrate and sulphate assimilation. (B) Selected categories of regulatory genes. This shows in more detail the condensed area in Fig 6C.

Supplemental Figure S5. AKIN10 deregulated transcripts were extracted from Baena-Gonzalez et al (2007). Transcript levels were normalized on the level at the end of the night in the same experiment, and kmeans-clustered. The panels show clusters of transcripts that are induced o rrepressed in AKIN10 overexpressing plants. The numbers of transcripts in each cluster are indicated in the panels.