Widening the landscape of transcriptional regulation of algal photoprotection

Authors

Marius Arend[1,2†], Yizhong Yuan[3†], M. Águila Ruiz-Sola[3‡], Nooshin Omranian[1,2,4], Zoran Nikoloski[1,2,4*], Dimitris Petroutsos[3 *]

Affiliations

[1]Bioinformatics Group, Institute of Biochemistry and Biology, University of Potsdam, 14476 Potsdam, Germany

[2]Systems Biology and Mathematical Modeling Group, Max-Planck-Institute of Molecular Plant Physiology, 14476 Potsdam, Germany

[3]University of Grenoble Alpes, CNRS, CEA, INRAE, IRIG-LPCV, 38000 Grenoble, France

[4]Bioinformatics and Mathematical Modeling Department, Center of Plant Systems Biology and Biotechnology, 4000 Plovdiv, Bulgaria

[†] these authors contributed equally to this work

[‡] Current address: Instituto de Bioquímica Vegetal y Fotosíntesis, Universidad de Sevilla-CSIC, 41092, Sevilla, SPAIN

[*]**Corresponding Authors**: nikoloski@mpimp-golm.mpg.de (Z.N.), dimitris.petroutsos@cea.fr (D.P.)

Abstract

Availability of light and CO_2, substrates of microalgae photosynthesis, is frequently far from optimal. Microalgae activate photoprotection under strong light, to prevent oxidative damage, and the CO_2 Concentrating Mechanism (CCM) under low CO_2, to raise intracellular CO_2 levels. The two processes are interconnected; yet, the underlying transcriptional regulators remain largely unknown. By employing a large transcriptomics data compendium of *Chlamydomonas reinhardtii's* responses to different light and carbon supply we reconstructed a consensus genome-scale gene regulatory network from complementary inference approaches and used it to elucidate the transcriptional regulation of photoprotection. We showed that the CCM regulator LCR1 also controls photoprotection, and that QER7, a Squamosa Binding Protein, suppresses photoprotection- and CCM-gene expression under the control of the blue light photoreceptor Phototropin. Along with demonstrating the existence of regulatory hubs that channel light- and CO_2-mediated signals into a common response, our study provides a unique resource to dissect gene expression regulation in this microalga.

37 **Introduction**

38 Photosynthetic microalgae convert light into chemical energy in the form of ATP and NADPH to fuel the CO_2
39 fixation in the Calvin–Benson cycle[1]. They have evolved to cope with rapid fluctuations in light[2] and inorganic
40 carbon[3] availability in their native habitats. When absorbed light exceeds the CO_2 assimilation capacity, the
41 formation of harmful reactive oxygen species can lead to severe cell damage; this is prevented by the
42 activation of photoprotective mechanisms, collectively called non-photochemical quenching (NPQ). NPQ
43 encompasses several processes that are distinguished in terms of their timescales[2], among which the rapidly
44 reversible energy-quenching (qE) is, under most circumstances, the predominant NPQ component[2,4]. The
45 major molecular effector of qE in Chlamydomonas is the LIGHT HARVESTING COMPLEX STRESS RELATED
46 protein LHCSR3, encoded by the *LHCSR3.1* and *LHCSR3.2* genes[5] that slightly differ only in their promoters;
47 LHCSR1 can also contribute significantly in qE under conditions where LHCSR3 is not expressed[6,7]. PSBS, the
48 key qE effector protein in higher plants[8] is encoded as two highly similar paralogues *PSBS1* and *PSBS2* in
49 *Chlamydomonas*[9]. They are only transiently expressed in *Chlamydomonas* under high light (HL)[9,10] and
50 accumulate under UV-B irradiation[11]; their precise contribution in *Chlamydomonas* photoprotective
51 responses is still unresolved[12].

52 Intracellular levels of CO_2 are modulated by the availability of its gaseous and hydrated forms[3] in the culture
53 media and the supply of acetate, that is partly metabolized into CO_2[7,13]. Under low CO_2, *Chlamydomonas*
54 activates the CO_2-concentrating mechanism (CCM) to avoid substrate-limitation of photosynthesis by raising
55 the CO_2 concentration at the site of RuBisCO, where CO_2 is assimilated[3]. The CCM mainly comprises of
56 carbonic anhydrases (CAHs) and of inorganic carbon transporters. Almost all CCM-related genes are under
57 the control of the nucleus-localized zinc-finger type nuclear factor CIA5 (aka CCM1)[14-16], including the Myb
58 Transcription Factor LOW-CO_2 -STRESS RESPONSE 1 (LCR1) that controls expression of genes coding for the
59 periplasmic CAH1, the plasma membrane-localized bicarbonate transporter LOW CO_2-INDUCED 1 (LCI1), and
60 the low-CO_2 responsive LCI6, whose role remains to be elucidated[17]. CIA5 is also a major qE regulator
61 activating transcription of genes encoding LHCSR3 and PSBS, while repressing accumulation of LHCSR1
62 protein[7].

63 LHCSR3 expression relies on blue light perception by the photoreceptor phototropin (PHOT)[18], on calcium
64 signaling, mediated by the calcium sensor CAS[19] and on active photosynthetic electron flow[18-20], likely via
65 indirectly impacting CO_2 availability[7]. The critical importance of CO_2 in LHCSR3 expression is demonstrated by
66 the fact that changes in CO_2 concentration can trigger LHCSR3 expression[21-23] even in the absence of light[7].
67 Accumulation of *LHCSR1* and *PSBS* mRNA is under control of the UV-B photoreceptor UVR8[11] and PHOT[24,25]

68 and is photosynthesis-independent[20,25]. While LHCSR1 is CO_2/CIA5 independent at the transcript level[7,25], PSBS
69 is responsive to CO_2 abundance and is under partial control of CIA5[7]. A Cullin (CUL4) dependent E3-ligase[24,26,27]
70 has been demonstrated to post-translationally regulate the transcription factor (TF) complex of CONSTANS
71 (CrCO)[26] and NF-Y isomers[27], which bind to DNA to regulate the transcription of *LHCSR1*, *LHCSR3*, and *PSBS*.
72 The putative TF and diurnal timekeeper RHYTHM OF CHLOROPLAST 75 (ROC75) was shown to repress LHCSR3
73 under illumination with red light[28].

74 Here, we employed a large compendium of RNAseq data from *Chlamydomonas* to build a gene regulatory
75 network (GRN) underlying light and carbon responses, and thus reveal the transcriptional regulation of qE at
76 the interface of these responses. The successful usage of RNAseq data to infer GRNs has been demonstrated
77 in many studies[29-31], although the data pose some challenges that require careful consideration. All of the
78 developed approaches to infer a GRN quantify the interdependence between the transcript levels of TF-
79 coding genes and their putative targets; the resulting prediction model serves as a proxy for the regulatory
80 strength that the product of the TF-coding gene exerts on its target(s). It is usually the case that the number
81 of observations (samples) is considerably smaller than the number of TFs used as predictors, leading to
82 collinearity of the transcript levels and associated computational instabilities; further, as an artifact of the
83 computational techniques, some of the inferred regulations may be spurious[32-34]. To address these issues,
84 here we took advantage of combining the outcome of multiple regularization techniques and post-processing
85 to increase the robustness of identified interactions[33-35]. In contrast to our approach, the existing predicted
86 GRNs of *Chlamydomonas* either focused on nitrogen starvation[36] or used a broad RNAseq data compendium,
87 not tailored to inferring regulatory interactions underlying responses to particular cues[37]. Moreover, these
88 GRNs were not obtained by combining the outcomes from multiple inference approaches, shown to increase
89 accuracy of predictions[30], and their quality was not gauged against existing knowledge of gene regulatory
90 interactions.

91 We used an RNAseq data compendium of 158 samples (**Supplementary Table 1**) from *Chlamydomonas*
92 cultures exposed to different light and carbon supply as input to seven benchmarked GRN inference
93 approaches that employ complementary inference strategies[29,30]. We assessed the performance of each
94 approach based on a set of curated TF-target gene interactions with experimental evidence from
95 *Chlamydomonas.* Based on this assessment, we integrated the outcome of the five best performing
96 approaches into a unique resource, a consensus network of *Chlamydomonas* light- and carbon-dependent
97 transcriptional regulation. We used the consensus network to reveal regulators of qE genes and
98 demonstrated the quality of predictions by validating two of the six tested candidates. We showed that LCR1

99 acts as a qE regulator by activating the expression of LHCSR3, and demonstrated that qE-REGULATOR 7 (QER7)
100 is a repressor of qE and CCM-gene expression. Our work consolidates the extensive co-regulation of CCM and
101 photoprotection[7] based on the untargeted assessment of the obtained genome-scale GRN.

102 Results

103 Computationally inferred GRN recovered known regulatory interactions underlying qE and CCM in
104 *Chlamydomonas*

105 We first aimed to employ published and here generated RNAseq data sets capturing the transcriptional
106 responses of *Chlamydomonas* to light and acetate availability to infer the underlying GRN. To this end, we
107 obtained data from two publicly available transcriptomics studies of synchronized chemostat wild-type (WT)
108 cultures grown in a 12h/12h light dark scheme and sampled in 30 min to 2h intervals[38,39]. We combined these
109 with our RNAseq data generated from mixotrophically or autotrophically grown batch cultures of the WT and
110 *phot* mutant acclimated to low light (LL) or exposed to HL (**Methods, Supplementary Table 1**). These data
111 sets capture the expected expression patterns of the key genes involved on CCM and qE (**Fig. 1a**) in response
112 to changes in acetate availability and light intensity. Specifically, we found strong up-regulation of these genes
113 in the light[7,25], and a marked inhibition of *LHCSR3.1/2* and CCM genes by acetate as previously described[7,40].

114 We employed these data together with a list of 407 transcription factors from protein homology studies[41,42]
115 **(Methods, Supplementary Table 2)**, as input to seven GRN inference approaches to robustly predict TF-target
116 interactions, as shown in benchmark studies[30]. Since there exists no study that experimentally probes TF
117 binding to DNA on a genome-scale, we first curated a list of known, experimentally validated regulatory
118 interactions underlying CCM and qE in *Chlamydomonas*[22,26-28], to assess the quality of GRNs inferred by the
119 different approaches (**Fig. 1b**). As a negative control, we considered the lack of effect of the SINGLET OXYGEN
120 RESISTANT 1 (SOR1) TF on *PSBS1* transcript levels in diurnal culture[43]. The comparative analysis of the
121 predicted and known interactions demonstrated that two of the applied approaches for GRN inference (i.e.
122 ARACNE[44] and global silencing[45]) are unable to recover any literature interactions when using a network
123 density threshold of 10% of all possible TF-TF and TF-target interactions. This poor performance is due either
124 to over-trimming or issues with the validity of the underlying assumptions, as seen in other case studies[46].
125 We therefore considered only the remaining five approaches, namely: Graphical Gaussian Models (GGM),
126 Context Likelihood of Relatedness (CLR), Elastic Net regression, Gene Network Inference with Ensemble of
127 Trees (GENIE3), and Network Deconvolution to infer a consensus GRN by using the Borda count election
128 method[30,47] (**Methods**). We then ranked the inferred interactions within each approach and quantified the

129 variability of ranks for the known TF-target gene interactions. We found that the average standard deviation
130 of the ranks of the TF-target gene interactions within an approach is larger than the average standard
131 deviation for the rank of a TF-target gene interaction across the five approaches (**Extended Data Fig. 1**, **Fig.**
132 **1b**). This observation suggested that the properties of a given TF-target gene interaction have a stronger
133 influence on its assigned rank than the inference approach used. More specifically, we noted that the
134 regulation of LHCSR genes by the two NF-Y paralogues and the induction of LCR1 by CIA5 are not recovered
135 by any of the used approaches; this is in line with reports showing that CIA5 is constitutively expressed and
136 regulated post-translationally[16]-not reflected in the transcriptomics data. Further, NF-Y factors that rely on
137 complex formation with CrCO to regulate their targets[27], may also act via unresolved posttranslational
138 mechanisms. Importantly, the regulatory interactions of the CCM effector genes *LCI1* and *CAH1* by *LCR1* are
139 assigned very high ranks (top 1%) by the employed approaches (except for GGM) considered in the consensus
140 GRN; moreover, the experimentally falsified interaction of *SOR1* and *PSBS1* transcripts[43] is correctly discarded
141 by all approaches (**Fig. 1b**). In addition, we observed that CLR and GENIE3 demonstrated the best performance
142 with respect to the set of known interactions. For instance, they identified the regulation of *LHCSR3.1* by
143 CrCO[26,27] and of *PSBS1* by NF-YB[27] (**Fig. 1b**). Generalization of this ranking beyond the known interactions
144 underlying qE and CCM processes is challenging, due to the lack of genome-scale gold standard, and we
145 therefore opted to combine the results of the five approaches, that showed comparable performance, in the
146 consensus GRN **(Methods, Supplementary Table 3)** to increase robustness of the predictions. Our analyses
147 of the overlap between the consensus and individual GRNs and the enrichment of TF-TF interactions
148 demonstrated the robustness of the inferred interactions **(Supplementary Note 1, Extended Data Fig. 2)**.

149 **Consensus GRN pinpoints LCR1 as a regulator of qE-related genes**

150 Using the consensus GRN, we inferred direct regulators of *LHCSR* and *PSBS* genes) and ranked them according
151 to the score resulting from the Borda method (**Methods**)[30,47]. Mutants were available for four of the top ten
152 of TFs with strongest cumulative regulatory effect on qE-related genes (**Fig. 2a, Supplementary Table 4):** Two
153 knock-out mutants of previously uncharacterized genes were ordered from the CliP library[48], which we
154 termed *qE-regulators 4* and *6* (*qer4*, *qer6*; see **Extended Data Fig. 3** for the genotyping of these mutants).
155 Additionally, we obtained an over-expressor line of the N-acetyltransferase LCI8[49] and the knock-out strain of
156 the known CCM regulator LCR1[17]. We tested for a regulatory effect by switching LL-acclimated mutant strains
157 and their respective WT background to HL for 1h and quantified transcript levels of qE-related genes. For *qer4*,
158 *qer6*, and *lci8-oe* we could not observe an effect on the transcript levels of investigated genes after HL
159 exposure (**Extended Data Fig. 4**). Thus, qer4 and qer6 are considered false positive predictions of the GRN,

160 despite the fact that *qer4* accumulated 1.5 times more *LHCSR3.1* under LL than the WT. A review of the closest
161 orthologs of LCI8 together with the experimental data indicate that it is likely involved in arginine synthesis[49]
162 and wrongly included as histone acetylase in the list of TFs. Interestingly, LCR1, the highest ranking among
163 the tested regulators showed significantly decreased expression of LHCSR3 at both the gene (3 times lower,
164 **Fig. 2b**) and protein level (4 times lower, **Fig. 2c,d**) compared to the WT; as a result, *lcr1* developed very low
165 NPQ and qE **(Fig. 2e)**. Complementation of *lcr1* with the knocked-out gene (strain *lcr1-C*) restored LHCSR3
166 gene and protein expression as well as the qE phenotype (**Fig. 2b-e**). Interestingly, the *lcr1* mutant over-
167 accumulated LHCSR1 and PSBS both at the transcript and at the protein level (**Fig. 2b-d**); Complementation
168 of *lcr1* with the knocked-out gene (strain *lcr1-C*) restored LHCSR3 gene and protein expression as well as the
169 qE phenotype (**Fig. 2b-e**). Because pre-acclimation conditions impact qE gene expression[25] we conducted
170 independent experiments in which cells were acclimated to darkness before exposure to HL and we obtained
171 very similar results. Our data demonstrated that *lcr1* showed significantly lower expression of LHCSR3 and
172 higher expression of LHCSR1/PSBS at both the gene (**Extended Data Fig. 5a**) and protein level (**Extended Data**
173 **Fig. 5b,c**), and had lower NPQ and qE (**Extended Data Fig. 5d**) than the WT, although the higher expression
174 levels of *LHCSR1* gene were not rescued by the complementation with the missing *LCR1* gene (**Extended Data**
175 **Fig. 5a**). Altogether, our data show that LCR1 is a regulator of qE by activating *LHCSR3.1* transcription and
176 repressing LHCSR1 and PSBS accumulation.

177 Further, we revisited the role of LCR1 in regulating CCM genes[17] by analyzing expression of selected CCM
178 genes in WT, *lcr1* and *lcr1-C* cells shifted from LL or darkness to HL, conditions favoring CCM gene expression[7].
179 We first confirmed that under our experimental conditions *lcr1* could not fully induce *LCI1* (**Extended Data**
180 **Fig. 6a, b**) in accordance to the report of the discovery of LCR1[17]. Our analyses further showed a statistically
181 significant impairment of *lcr1* in inducing genes encoding the Ci transporters LOW-CO_2-INDUCIBLE PROTEIN
182 A (LCIA), HIGH-LIGHT ACTIVATED 3 (HLA3), CHLOROPLAST CARRIER PROTEIN 1 (CCP1) and BESTROPHINE-LIKE
183 PROTEIN 1 (BST1) as well as the carbonic anhydrase CAH4, when shifted from LL or dark to HL (**Extended Data**
184 **Fig. 6a, b**), indicating that the role of LCR1 in low-CO_2 gene expression extends beyond the regulation of gene
185 expression of *CAH1*, *LCI1* and *LCI6*[17].

186 **PHOT-specific GRN reveals a novel repressor of qE and CCM**

187 The light-dependent induction of LHCSR3 is predominantly mediated by the blue light photoreceptor PHOT[18].
188 To analyze the PHOT-dependent transcriptional regulators, we first identified the interactions shared by the
189 consensus network and a GRN inferred only by GENIE3, due to its good performance, using the RNAseq data
190 from samples of *phot* and WT acclimated to LL and after 1h exposure to HL (**Methods**). This so-called PHOT-

specific GRN (**Methods**, **Supplementary Table 5**) comprises regulatory interactions underlying the transcriptomic changes observed in the *phot* mutant while burrowing the statistical power of the RNAseq compendium.

When investigating the top 10 regulators of qE genes in this PHOT GRN the interactions where weighted based on the importance score from GENIE3 (**Supplementary Table 6, Fig. 3a**). Among these interactions we recovered the previously discussed false positive LCI8 gene (see above, **Fig. 3a**) (**Extended Data Fig. 4**). In addition, two previously reported regulators of qE, ROC75[28] and CrCO[26,27], are also included in this list of PHOT-dependent regulators. These observations are in line with an existing hypothesis[24] suggesting that a CUL4-dependent E3-ligase targeting CrCO[,26] acts downstream of PHOT. ROC75 has been previously reported to act independently of the PHOT signal based on qPCR studies of the mutant grown synchronously under different light spectra[28]. In our RNAseq data, gathered under continuous white light, we observed a significant difference in expression levels of ROC75 between WT and *phot* (log2 fold-change = 1.03, adj. p-value = $1.80*10^{-7}$).

The fact that several regulators showed larger regulatory strength than CrCO in the PHOT GRN indicates the existence of yet unreported regulators of qE effector genes in the PHOT signaling pathway. This is in line with existing results[26], showing that the knock-out of CrCO is insufficient to fully abolish light-dependent activation of LHCSR3. Following this reasoning we obtained *qer1* and *qer7*, the available regulator candidates mutants, from the CLiP library[48] , (for genotyping see **Extended Data Fig. 3**). Our results show higher mRNA and protein levels of LHCSR1 in the *qer1* mutant (**Extended Data Fig. 7**); however, this could not be rescued by ectopic expression of the *QER1* gene in the *qer1* mutant background (**Extended Data Fig. 7**). We found significant upregulation of *LHCSR3.1* gene expression in the *qer7* mutant (1.7 times, **Extended Data Fig. 8a**) also reflected in higher NPQ (**Extended Data Fig. 8b)** and qE levels (**Extended Data Fig. 8c**) which we followed up in more detail. To this end, we ectopically expressed the WT *QER7* gene in the *qer7* mutant and generated the complemented strain *qer7-C* that expressed *QER7* to levels similar to those WT (**Extended Data Fig. 3c**). As a result, the *qer7-C* strain showed reduced *LHCSR3* gene expression, NPQ and qE levels as compared with the *qer7* mutant (**Extended Data Fig. 8a-c**). *LHCSR1* and *PSBS* seemed to be unaffected in the qer7 in these LL to HL transition experiments (**Extended Data Fig. 8a**). As with LCR1, we also performed dark to HL experiments to further characterize the photoprotective responses of *qer7*; under these conditions, *qer7* accumulated significantly more *LHCSR1* (1.7 times) and *PSBS1* (2.2 times) while *LHCSR3* remained unaffected (**Extended Data Fig. 8d**). As in the LL to HL experiments (**Extended Data Fig. 8b, c**) *qer7* showed more NPQ and qE (**Extended Data Fig. 8e, f**). Complementation of *qer7* with the missing *QER7* gene restored all phenotypes

222 (*LHCSR1*, *PSBS*, NPQ, qE; **Extended Data Fig. 8d-f**). These data validate the prediction of QER7 as regulator of
223 qE gene expression (**Fig. 3a**) and indicate that QER7 regulates different subsets of qE genes depending on the
224 pre-acclimation conditions; *LHCSR3* when preacclimation is under LL, *LHCSR1* and *PSBS* when pre-acclimation
225 occurs in darkness. Motivated by these findings and given the fact that most of the Chlamydomonas
226 transcriptome, undergoes diurnal changes according to biological function[39] we decided to address the role
227 of QER7 in regulating qE genes under light/dark cycles. We synchronized WT, *qer7* and *qer7-C* cells in 12h
228 L/12h D cycle and exposed them to HL right after the end of the dark phase. Our results revealed that under
229 these conditions QER7 functions as a repressor of all qE-related genes; the *qer7* mutant expresses significantly
230 higher LHCSR3, LHCSR1 and PSBS both at the gene (**Fig. 3b**) and protein (**Fig. 3c,d**) level, and exhibits higher
231 NPQ and qE (**Fig. 3e**), with all phenotypes rescued in the *qer7-C* complemented line. Previous protein
232 homology studies identified QER7 as Squamosa Binding Protein[50] or bZIP TF[51], and here, we provide the first
233 functional annotation of QER7 as a novel qE regulator.

234 **QER7 co-regulates qE-related and CCM genes**

235 The recent findings that the regulatory role of CIA5[7] and LCR1 (**Fig. 2 and Extended Data Fig. 5**) extends
236 beyond CCM to also control qE-related gene expression, prompted us to also inspect the expression levels of
237 CCM genes in synchronized *qer7* cells (**Fig. 4a**). Indeed, for four of these transcripts (*CAH4*, *LCIA*, *CCP1*, *HLA3*)
238 we observed a significant upregulation in *qer7* after HL exposure that was reversed by complementation with
239 the *QER7* gene, indicating that QER7 supresses expression of CCM genes; the suppression role of QER7 on
240 CCM genes was only observable under HL, conditions that favor CCM gene expression[7] and not under LL (**Fig.
241 4a**). Interestingly, and in line with QER7 acting as suppressor of CCM gene expression (**Fig. 4a**), we found LCR1
242 and QER7 among the top 10 regulators of the CCM genes probed by qPCR in the inferred GRNs (**Extended
243 Data Fig. 9**).

244 After observing the inhibition of CCM gene transcription by QER7, we investigated the signaling pathway
245 upstream of QER7. To this end, we quantified *QER7* gene expression in synchronized *phot* cultures. As
246 expected, we observe a strong effect of the *PHOT* knock-out showing that PHOT suppresses *QER7* expression
247 (**Fig. 4b**) implying that blue-light perception via PHOT is involved in the regulation of CCM related gene
248 expression. Interestingly, *phot* expressed *LCR1* to levels similar to those in WT (**Fig. 4c**). Thus, while sharing
249 part of their target genes, the two TFs mediate different signals. These data demonstrate that the PHOT-
250 specific GRN successfully pinpoints TFs that depend on PHOT activity while retaining the statistical power to
251 recover genuine regulatory interactions.

252 The two qE regulators that we validated in this study also regulate CCM genes. Therefore, we next
253 investigated to what extent the observed coregulation pattern applies to the global, known transcriptional
254 regulation of low CO_2 and light stress responsive genes. To this end, we took advantage of the size of the
255 presented genome-scale GRNs and compiled a list of genes putatively involved in photoprotection
256 (**Supplementary Table 7**) or the CCM (**Supplementary Table 8**); we then extracted the 10 TFs exhibiting the
257 strongest regulatory strength on the genes in the compiled lists. We found six (empirical p-value<0.001,
258 **Methods**) and four (empirical p-value < 0.01) of the top 10 regulators to be shared between these two
259 responses in the consensus **(Fig. 5a,b**) and the PHOT-specific GRN, respectively (**Figure 5c,d**). The significant,
260 large number of shared regulators is a strong indication that co-regulation of photoprotective and carbon
261 assimilatory processes is a principal feature of Chlamydomonas' transcriptional regulatory program.

262 ## Discussion

263 The molecular actors and structure of the transcriptional regulatory mechanisms that shape Chlamydomonas'
264 response to differential light and carbon availability are largely unknown, although they are paramount to
265 survival of Chlamydomonas and offer valuable targets for biological engineering. Here we set out to elucidate
266 the GRN underlying the response to light and carbon availability by combining the results from five
267 complementary inference approaches and data from 158 RNAseq samples of cultures responding to these
268 cues. The obtained GRN enabled us to identify novel regulators controlling qE. Experimentally testing
269 available mutant strains for six of these candidates we were able to validate two TFs: QER7, suppressing qE-
270 and CCM-gene expression, and LCR1, activating LHCSR3 and suppressing LHCSR1 and PSBS expression. The
271 two TFs belong to independent signalling pathways: LCR1 has been known to controlled by CIA5 while we
272 found that QER7 is controlled by Phototropin; this last finding adds a new layer of regulation of CCM by blue
273 light sensing. We complemented the finding of the involvement of LCR1 in the regulation of qE related genes
274 with an analysis of the genome-scale co-regulation of CCM and photoprotective genes. Indeed, we observed
275 significant number (six out of the top 10) of regulators targeting both processes in the inferred consensus
276 GRN. Together with previous studies, our results provide evidence for extensive signaling crosstalk between
277 light and carbon sensing and name a set of TFs that likely integrate these disparate signals into a common
278 transcriptional response. The GRN gives an unbiased representation of the genome-scale regulators acting in
279 light and carbon sensing. Thus, it represents a powerful resource for future dissection of the transcriptional
280 regulation of responses of *Chlamydomonas* to light and carbon availability.

281

282 **Methods**

283 **Transcriptome analysis**

284 We assembled a compendium of RNAseq data (**Supplementary Table 1**) that capture regulation of light-
285 dependent processes by combining in-house produced RNAseq measurements with publicly available data
286 from two studies of densely sampled diurnal cultures of Chlamydomonas[38,39]. For the samples in the acetate
287 time-resolved experiment, adapter sequences were specifically trimmed from raw reads using BBduk [52]
288 (ktrim=r k=30 mink=12 minlen=50). Raw reads of the diurnal transcriptome study from Strenkert et al.[39] were
289 obtained from NCBI GEO database (GSE112394). Reads were aligned to the Chlamydomonas reference
290 transcriptome[53] available from JGI Phytozome (Assembly version 5) using RNA STAR aligner. The BAM files
291 obtained from these measurements were analyzed using HTSeq-count[54] (standed=reverse) to create raw read
292 count files. The raw read counts from Zones et al. [38] were obtained as .tsv from NCBI GEO (GSE71469). The
293 final data set consists of 158 samples from 62 experimental conditions or time points (**Supplementary Table**
294 **1**). Genes with less than 1 count per million in at least 9 measurements where discarded and the remainder
295 were voom[53] transformed and normalized using library normalization factors based on the TMM[55] approach
296 as implemented in the R Bioconductor package edgeR[56].

297 **Transcription factor set from comparative genomics**

298 To reduce the set of parameters in our network model, we compiled transcription factor (TF) annotations for
299 the Chlamydomonas genome based on proteome homology studies. We obtained the proteomes and protein
300 IDs of predicted Chlamydomonas TFs from Pérez-Rodríguez et al.[41] Since these predictions were built based
301 on the older Chlamydomonas assembly, we first used the conversion table provided by Phytozome to convert
302 JGI4 to Crev5.6 IDs. For the TFs that could not be recovered by this approach we used the Phytozome BLAST
303 tool to align these sequences against the Crev5.6 proteome (BLASTP, E threshold: -1, comparison matrix:
304 BLOSUM62, word length: 11, number of reported alignments: 5). The reported hits were filtered for sequence
305 identity > 97% and gaps ≤ 1 . If sequences mapped multiple times to the same Crev5.6 gene ID, only the hit
306 alignment closest to the N-terminus of the query sequence was kept. The hit was only accepted if the
307 alignment started at least six residues from the N-terminus of the hit sequence. For Crev5.6 loci that had
308 multiple JGI4 TF queries assigned to them the best hit was selected manually. This set was then extended by
309 the TFs found in the study of Jin et al.[42] and the regulators in the manually curated set of CCM and qE
310 regulatory interactions (**Supplementary Tables 7 and 8**). Using this procedure, we compiled a list of 407
311 Chlamydomonas TFs (**Supplementary Table 2**) to be considered as regulators in the inferred networks.

312 **Gene regulatory network inference**

313 The CLR and ARACNE approach were based on all replicate measurements; for all other inference methods
314 the median from each condition was used as input. All input matrices where standardized gene-wise. If not
315 explicitly stated in the respective paragraph the implementations of all GRN inference approaches were
316 applied with their default settings.

317 ***Graphical Gaussian Models***

318 The network inferred from a Graphical Gaussian model of gene regulation was obtained using the
319 implementation of the partial correlation estimate from Schäfer et al.[34] as implemented in the R GeneNet
320 package. All interactions between TFs and another gene/TF with non-zero partial correlations were included
321 as network edges.

322 ***GENIE3***

323 The random forest-based network from GENIE3 was generated using the R Bioconductor implementation
324 provided by the authors[57]. We used only expression levels of TFs as predictors.

325 ***Elastic net regression***

326 A linear regression based network was obtained using the elastic net algorithm[35]. A model was fit for each
327 gene using the expression levels of all TFs as predictors. The two hyperparameters $\lambda 2$ (quadratic penalty) and
328 s (fraction of L1 norm coefficients) were tuned for each gene model using 6-fold cross validation. The 2D
329 parameter space scanned was $\lambda 2$={0,0.001,0.01,0.05,0.1,0.5,1,1.5,2,10,100} and
330 s={0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9}. The R2 value for each model was calculated as

$$R^2 = 1 - \frac{\sum(y - \hat{y})}{var(y)} \quad (1)$$

331 with y marking the vector of observed expression values and $\hat{y}$ the model predictions. Models with a negative
332 $R2$ value were discarded as regularization artifacts. The results of the remaining models were assembled into
333 a network in which interactions were ranked by regression coefficients β normalized by the maximum
334 absolute coefficient.

$$\beta_n = \frac{\beta}{\max|\beta|} \quad (2)$$

335

336 ***CLR and ARACNE***

The implementation of mutual information (MI)-based network inference approaches from the R package minet[58] was used. Pairwise MI was estimated based on the Spearman correlation as proposed by Olsen et al.[59]. Two networks were constructed based on these MI estimates. Using the CLR approach[60] non-significant interactions were removed based on the z-scores calculated from the marginal distributions of MI values for each gene pair. Alternatively, the ARACNE algorithm[44] was used to prune the network based on the data processing inequality. For both networks only interactions originating from a TF were taken into consideration and edges were ranked according to the assigned MI value.

Deconvolution and Silencing

For the two networks based on decomposition of the interaction matrix *G* the Pearson correlation matrix obtained from gene expression values was used as input.

The deconvolution approach introduced by Feizi et al. [61] was implemented as previously described[46]. The eigenvalue scaling factor β was initialized as β=0.9 and iteratively reduced in increments of 0.05 until the largest eigenvector of the direct interaction matrix generated by deconvolution was smaller than 1. Edges were ranked according to the deconvoluted interaction matrix.

The Silencing approach as described by Barzel et al.[45] was implemented in R. The proposed approximation of the direct interaction matrix *S* in which spurious interactions are silenced relies on the invers of the observed correlation matrix *G*. In our implementation we used the Moore-Penrose pseudoinverse in case *G* was close to singular. In the resulting network edges were ranked according to the approximated silenced interaction matrix.

Consensus network construction

To improve network quality[30], we built a consensus network integrating the GRN models inferred by the different approaches introduced above. To this end, we used the Borda count election method[47] whereby the rank *r* of an interaction *I* in the consensus network built on the predictions from *k* approaches is given by arithmetic mean of the ranks in the individual networks

$$r_{consens}(I) = \frac{\sum_{i=1}^{k} r_i(I)}{k}. \quad (3)$$

Following the reasoning of Feizi et al. [61]) in this integration only the top 10% all possible edges in the GRN (625815) were considered from each individual ranking. For an edge that was not assigned a rank by some approaches, the missing ranks were set to 10% of all possible edges plus one.

364 Using this integration method, we assembled a consensus network based on all approaches to compare
365 predictions from all GRN inference approaches (**Extended Data Fig. 1**). Due to this comparison and inability
366 to recover known interactions (**Fig. 1b**) the rankings derived from ARACNE and Silencing were only considered
367 in **Extended Data Fig. 1** and excluded from the final consensus network used for all other analyses. As with
368 the individual networks returned by the different approaches the consensus network **(Supplemental Table 3)**
369 was trimmed to the top 10% of all possible edges according to the integrated ranks . The weight of edges in
370 the final network was set as $r_{consens}^{-1}$.

371 **PHOT-specific network**

372 To investigate the PHOT-specific regulatory interactions genes that are differentially expressed between phot
373 mutant and wt under low and high light were inferred. To this end, transcript counts of genes with more than
374 1 count per million in at least four replicates from these conditions (**Supplementary Table 1**) were tested for
375 differential expression using the R packages limma[62], DeSeq2[63], and edgeR[56]. Only genes deemed significant
376 by all three tools after Benjamini-Hochberg correction for a false discovery rate of 0.05 were considered
377 differentially expressed with respect to PHOT mutation

378 In the next step, we focused on the normalized and scaled expression levels from these differentially
379 expressed genes and the previously mentioned conditions, to infer a PHOT-specific GRN using GENIE3. To
380 improve robustness of this network, which was obtained from a comparably sparse data set, we only
381 considered the edges in the intersection with the final consensus network. Again, for both networks only the
382 top 10% of possible edges were taken into account. Therefore, the obtained PHOT-network represents a
383 subnetwork of the final consensus in which edges are weighted by the "PHOT specific" GENIE3 importance
384 measure **(Supplementary Table 5)**.

385 **Identification of major regulators**

386 We compiled a manually curated list of possible target genes known to be involved in the processes of qE
387 (*LHCSR1, LHCSR3.1/2, PSBS1/2*), photoprotection (**Supplementary Table 7**), and CCM (**Supplementary Table**
388 **8**). Based on the assumption that major regulators act on several genes important for a biological process,
389 the regulatory strength of a candidate regulator (for the given process) was determined by the sum of edge
390 weights w_{ij} between this regulator and the k genes in the respective target gene set

$$C(TF) = \sum_{j=1}^{k} w_{TFj}. \quad (4)$$

391 **Empirical p-value calculation using Monte-Carlo simulation**

392 The one-sided p-value for the overlap between the regulators of CCM and photoprotective genes was
393 approximated by sampling the overlaps of random gene sets. To this end, we compiled two gene sets with
394 the same cardinality as the curated CCM and photoprotective genes. The genes in these sets where
395 randomly sampled without replacement from all targets in the respective networks. The 10 strongest
396 regulators of these two gene sets where then obtained as previously described and the overlap was
397 calculated as our sample statistic. This process was repeated 10,000 times and an empirical p-value was
398 calculated from the number of iterations, r, where the overlap was higher or equal to the observed value,
399 and the total number of iterations, n [64]:

$$p = \frac{r+1}{n+1}. \tag{5}$$

400 **Strains and conditions**

401 *C. reinhardtii* strains were grown under 15 µmol photons m^{-2} s^{-1}) in Tris-acetate-phosphate (TAP) media[65] at
402 23 °C in Erlenmeyer flasks shaken at 125 rpm. For all experiments cells were transferred to Sueoka's high salt
403 medium (HSM)[66] at 1 million cells mL^{-1} and exposed to light intensities as described in the text and figure
404 legends. For the investigation of the impact of acetate on the genome-wide transcriptome, HSM was
405 supplemented with 20 mM sodium acetate. *C. reinhardtii* strain CC-125 mt+ was used as WT. The *phot*
406 (depleted from *PHOT1*; gene ID: Cre03.g199000)*, was* previously generated[67] and recently characterized[25].
407 For synchronized cultures, the cells were grown in HSM for at least 5 days under a 12h light/12h dark cycle
408 (light intensity was set at 15 µmol photons m^{-2} s^{-1}; temperature was 18 °C in the dark and 23 °C in the light).
409 All CLiP mutant strains used in this study and their parental strain (CC-4533) were obtained from the CLiP
410 library (REF); *qer1* (LMJ.RY0402.072278), *qer4* (LMJ.RY0402.202963), *qer6* (LMJ.RY0402.162350), *qer7*
411 (LMJ.RY0402.118995). The *lcr1* (strain C44)*, lcr1-C* (strain C44-B7) and its parental strain Q30P3 as described
412 in[17] were a kind gift from Hideya Fukuzawa. Before performing phenotyping experiments, we first confirmed
413 that *lcr1* shows no expression of *LCR1* and that this is rescued in the *lcr1-C* strain (**Extended Data Fig. 10a**).
414 The *lci8* overexpressing line was purchased from the Chlamydomonas Resource center; strain CSI_FC1G01,
415 expressing pLM005-Cre02.g144800-Venus-3xFLAG in the CC-4533 background. Overexpression of LCI8-FLAG
416 was verified by immunoblotting against FLAG (**Extended Data Fig. 10b**).

417 To complement *qer1*, a 1152 bp genomic DNA fragment from *Chlamydomonas* CC-4533 was amplified by PCR
418 using KOD hot start DNA polymerase (Merck) and primers P11 and P12 (**Supplementary Table 9**). To
419 complement *qer7*, a 5755 bp fragment DNA fragment from *Chlamydomonas* CC-4533 was amplified by PCR

420 with Platinum superfii DNA Polymerase (Thermo Fisher Scientific) and primers P13 and P14 (**Supplementary**
421 **Table 9**). The PCR products were gel purified and cloned into pRAM118[68] by Gibson assembly[69] for expression
422 under control of the *PSAD* promoter. Junctions and insertion were sequenced and constructs were linearized
423 by EcoRV before transformation. Eleven ng/kb of linearized plasmid[70] mixed with 400 µL of 1.0 x10^7 cells mL^{-1}
424 were electroporated in a volume of 120 mL in a 2-mm-gap electro cuvette using a NEPA21 square-pulse
425 electroporator (NEPAGENE, Japan). The electroporation parameters were set as follows: Poring Pulse (300V;
426 8 ms length; 50 ms interval; one pulse; 40% decay rate; + Polarity), Transfer Pluse (20V; 50 ms length; 50 ms
427 interval; five pulses; 40% decay rate; +/- Polarity). Transformants were selected onto solid agar plates
428 containing 20 µg/ml hygromycin and screened for fluorescence by using a Tecan fluorescence microplate
429 reader (Tecan Group Ltd., Switzerland). Parameters used were as follows: YFP (excitation 515/12 nm and
430 emission 550/12 nm) and chlorophyll (excitation 440/9 nm and 680/20 nm). Transformants showing high
431 YFP/chlorophyll value were further analyzed by real time qPCR.

432 Unless otherwise stated, LL conditions corresponded to 15 µmol photons m^{-2} s^{-1} while HL conditions
433 corresponded to 300 µmol photons m^{-2} s^{-1} of white light (Neptune L.E.D., France; see[7] for light spectrum). All
434 experiments were repeated at least three times to verify their reproducibility.

435 **DNA Isolation and genotyping of CLiP mutants.**

436 Total genomic DNA from CLiP mutants and corresponding wild-type strain CC-4533 was extracted according
437 to the protocol suggested by CLiP website (https://www.chlamylibrary.org/). One µl of the extracted DNA
438 was used as a template for the PCR assays, using Phire Plant Direct PCR polymerase (Thermo Fisher
439 Scientific). To confirm the CIB1 insertion site in the CLiP mutants, gene-specific primers were used that
440 anneal upstream and downstream of the predicted insertion site of the cassette (primer pairs P3-P4, P7-P8,
441 P9-P10 and P5-P6 for *qer6*, *qer1*, *qer7* and *qer4* respectively*;* **Supplementary Table 9**). While all these
442 primers worked in DNA extracted from WT, they did not work in the DNA extracted from the mutants, with
443 the exception of *qer4* (**Extended Data Fig. 3**), therefore primers specific for the 5' and 3' ends of the CIB1
444 Cassette were additionally used. All the primers used for genotyping were shown in **Supplementary Table 9**.
445 We further confirmed the disruption of the genes of interest by quantifying their mRNA accumulation
446 (**Extended Data Fig. 3**).

447 **mRNA quantification**

448 Total RNA was extracted using the RNeasy Mini Kit (Qiagen) and treated with the RNase-Free DNase Set
449 (Qiagen). 1 µg total RNA was reverse transcribed with oligo dT using Sensifast cDNA Synthesis kit (Meridian

450 Bioscience, USA). qPCR reactions were performed and quantitated in a Bio-Rad CFX96 system using
451 SsoAdvanced Universal SYBR Green Supermix (BioRad). The primers (0.3 µM) used for qPCR are listed in
452 **Supplementary Table 10**. A gene encoding G protein subunit-like protein (GBLP)[71] was used as the
453 endogenous control , and relative expression values relative to *GBLP* were calculated from three biological
454 replicates, each of which contained three technical replicates. All primers using for qPCR (**Supplementary**
455 **Table 10**) were confirmed as having at least 90% amplification efficiency. In order to conform mRNA
456 accumulation data to the distributional assumptions of ANOVA, i.e. the residuals should be normally
457 distributed and variances should be equal among groups, Two-Way Analysis of Variance were computed with
458 log-transformed data Y=logX where X is mRNA accumulation[72].

459 **Immunoblotting**

460 Protein samples of whole cell extracts (0.5 µg chlorophyll or 10 µg protein) were loaded on 4-20% SDS-PAGE
461 gels (Mini-PROTEAN TGX Precast Protein Gels, Bio-Rad) and blotted onto nitrocellulose membranes. Antisera
462 against LHCSR1 (AS14 2819), LHCSR3 (AS14 2766), ATPB (AS05 085), CAH4/5 (AS11 1737) were from Agrisera
463 (Vännäs, Sweden); antiserum against PSBS was from ShineGene Molecular Biotech (Shanghai, China)
464 targeting the peptides described in Ref.[9]. ATPB was used as a loading control. An anti-rabbit horseradish
465 peroxidase–conjugated antiserum was used for detection. The blots were developed with ECL detection
466 reagent, and images of the blots were obtained using a CCD imager (ChemiDoc MP System, Bio-Rad). For the
467 densitometric quantification, data were normalized with ATPB.

468 **Fluorescence-based measurements**

469 Fluorescence-based photosynthetic parameters were measured with a a pulse modulated amplitude
470 fluorimeter (MAXI-IMAGING-PAM, HeinzWaltz GmbH, Germany). Prior to the onset of the measurements,
471 cells were acclimated to darkness for 15 min. Chlorophyll fluorescence was recorded during 10 min under 570
472 µmol m^{-2} s^{-1} of actinic blue light followed by finishing with 10 min of measurements of fluorescence relaxation
473 in the dark. A saturating pulse (200 msec) of blue light (6000 µmol photons m^{-2} sec^{-1}) was applied for
474 determination of Fm (the maximal fluorescence yield in dark-adapted state) or Fm' (maximal fluorescence in
475 any light-adapted state). NPQ was calculated as (Fm − Fm′)/Fm′ based on[73]; qE was estimated as the fraction
476 of NPQ that is rapidly inducible in the light and reversible in the dark.

477

478

References

1. Eberhard, S., Finazzi, G. & Wollman, F. A. The dynamics of photosynthesis. *Annu. Rev. Genet.* **42,** 463–515 (2008).
2. Erickson, E., Wakao, S. & Niyogi, K. K. Light stress and photoprotection in Chlamydomonas reinhardtii. *Plant J.* **82,** 449–465 (2015).
3. Moroney, J. V. & Ynalvez, R. A. Proposed carbon dioxide concentrating mechanism in Chlamydomonas reinhardtii. *Eukaryotic Cell* **6,** 1251–1259 (2007).
4. Niyogi, K. K. PHOTOPROTECTION REVISITED: Genetic and Molecular Approaches. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **50,** 333–359 (1999).
5. Peers, G. *et al.* An ancient light-harvesting protein is critical for the regulation of algal photosynthesis. *Nature* **462,** 518–521 (2009).
6. Dinc, E. *et al.* LHCSR1 induces a fast and reversible pH-dependent fluorescence quenching in LHCII in Chlamydomonas reinhardtii cells. *Proc. Natl. Acad. Sci. U.S.A.* **113,** 7673–7678 (2016).
7. Ruiz-Sola, M. *et al.* Photoprotection is regulated by light-independent CO2 availability. *bioRxiv* (2021). doi:10.1101/2021.10.23.465040
8. Li, X. P. *et al.* A pigment-binding protein essential for regulation of photosynthetic light harvesting. *Nature* **403,** 391–395 (2000).
9. Correa-Galvis, V. *et al.* Photosystem II Subunit PsbS Is Involved in the Induction of LHCSR Protein-dependent Energy Dissipation in Chlamydomonas reinhardtii. *J. Biol. Chem.* **291,** 17478–17487 (2016).
10. Tibiletti, T., Auroy, P., Peltier, G. & Caffarri, S. Chlamydomonas reinhardtii PsbS Protein Is Functional and Accumulates Rapidly and Transiently under High Light. *Plant Physiol.* **171,** 2717–2730 (2016).
11. Allorent, G. *et al.* UV-B photoreceptor-mediated protection of the photosynthetic machinery in Chlamydomonas reinhardtii. *Proc. Natl. Acad. Sci. U.S.A.* **113,** 14864–14869 (2016).
12. Redekop, P. *et al.* PsbS contributes to photoprotection in Chlamydomonas reinhardtii independently of energy dissipation. *BIOCHIMICA ET BIOPHYSICA ACTA-BIOENERGETICS* **1861,** 148183 (2020).
13. Hanawa, Y., Watanabe, M., Karatsu, Y., Fukuzawa, H. & Shiraiwa, Y. Induction of a high-CO2-inducible, periplasmic protein, H43, and its application as a high-CO2-responsive marker for study of the high-CO2-sensing mechanism in Chlamydomonas reinhardtii. *Plant Cell Physiol.* **48,** 299–309 (2007).
14. Moroney, J. V. *et al.* Isolation and Characterization of a Mutant of Chlamydomonas reinhardtii Deficient in the CO2 Concentrating Mechanism. *Plant Physiol.* **89,** 897–903 (1989).
15. Fukuzawa, H. *et al.* Ccm1, a regulatory gene controlling the induction of a carbon-concentrating mechanism in Chlamydomonas reinhardtii by sensing CO2 availability. *Proc. Natl. Acad. Sci. U.S.A.* **98,** 5347–5352 (2001).
16. Xiang, Y., Zhang, J. & Weeks, D. P. The Cia5 gene controls formation of the carbon concentrating mechanism in Chlamydomonas reinhardtii. *Proc. Natl. Acad. Sci. U.S.A.* **98,** 5341–5346 (2001).
17. Yoshioka, S. *et al.* The novel Myb transcription factor LCR1 regulates the CO2-responsive gene Cah1, encoding a periplasmic carbonic anhydrase in Chlamydomonas reinhardtii. *Plant Cell* **16,** 1466–1477 (2004).
18. Petroutsos, D. *et al.* A blue-light photoreceptor mediates the feedback regulation of photosynthesis. *Nature* **537,** 563–566 (2016).
19. Petroutsos, D. *et al.* The chloroplast calcium sensor CAS is required for photoacclimation in *Chlamydomonas reinhardtii*. *Plant Cell* **23,** 2950–2963 (2011).

20. Maruyama, S., Tokutsu, R. & Minagawa, J. Transcriptional regulation of the stress-responsive light harvesting complex genes in *Chlamydomonas reinhardtii*. *Plant Cell Physiol.* **55,** 1304–1310 (2014).
21. Miura, K. *et al.* Expression profiling-based identification of CO2-responsive genes regulated by CCM1 controlling a carbon-concentrating mechanism in Chlamydomonas reinhardtii. *Plant Physiol.* **135,** 1595–1607 (2004).
22. Fang, W. *et al.* Transcriptome-wide changes in Chlamydomonas reinhardtii gene expression regulated by carbon dioxide and the CO2-concentrating mechanism regulator CIA5/CCM1. *Plant Cell* **24,** 1876–1893 (2012).
23. Brueggeman, A. J. *et al.* Activation of the carbon concentrating mechanism by CO2 deprivation coincides with massive transcriptional restructuring in Chlamydomonas reinhardtii. *Plant Cell* **24,** 1860–1875 (2012).
24. Aihara, Y., Fujimura-Kamada, K., Yamasaki, T. & Minagawa, J. Algal photoprotection is regulated by the E3 ligase CUL4-DDB1DET1. *Nature Plants* **5,** 34–40 (2019).
25. Redekop, P. *et al.* Transcriptional regulation of photoprotection in dark-to-light transition- more than just a matter of excess light energy. *BioRxiv* (2021) https://doi.org/10.1101/2021.10.23.463292
26. Gabilly, S. T. *et al.* Regulation of photoprotection gene expression in Chlamydomonas by a putative E3 ubiquitin ligase complex and a homolog of CONSTANS. *Proc. Natl. Acad. Sci. U.S.A.* **116,** 17556–17562 (2019).
27. Tokutsu, R., Fujimura-Kamada, K., Matsuo, T., Yamasaki, T. & Minagawa, J. The CONSTANS flowering complex controls the protective response of photosynthesis in the green alga Chlamydomonas. *Nat Comms* **10,** 655–10 (2019).
28. Kamrani, Y. Y., Matsuo, T., Mittag, M. & Minagawa, J. ROC75 is an Attenuator for the Circadian Clock that Controls LHCSR3 Expression. *Plant Cell Physiol.* **59,** 2602–2607 (2018).
29. Omranian, N., Eloundou-Mbebi, J. M. O., Mueller-Roeber, B. & Nikoloski, Z. Gene regulatory network inference using fused LASSO on multiple data sets. *Sci Rep* **6,** 1–14 (2016).
30. Marbach, D. *et al.* Wisdom of crowds for robust gene network inference. *Nat. Methods* **9,** 796–804 (2012).
31. Fang, L. *et al.* GRNdb: Decoding the gene regulatory networks in diverse human and mouse conditions. *Nucl. Acids Res.* **49,** D97–D103 (2021).
32. Omranian, N. & Nikoloski, Z. in *Methods in Molecular Biology* **1629,** 283–295 (2017).
33. Huynh-Thu, V. A. & Sanguinetti, G. in *Methods in Molecular Biology* **1883,** 1–23 (2019).
34. Schäfer, J. & Strimmer, K. A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Stat Appl Genet Mol Biol* **4,** Article32 (2005).
35. Zou, H. & Hastie, T. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **67,** 301–320 (2005).
36. Gargouri, M. *et al.* Identification of regulatory network hubs that control lipid metabolism in Chlamydomonas reinhardtii. *J. Exp. Bot.* **66,** 4551–4566 (2015).
37. Lämmermann, N. *et al.* Ubiquitin ligase component LRS1 and transcription factor CrHy5 act as a light switch for photoprotection in <em>Chlamydomonas</em&gt. *bioRxiv* 2020.02.10.942334 (2020).
38. Zones, J. M., Blaby, I. K., Merchant, S. S. & Umen, J. G. High-Resolution Profiling of a Synchronized Diurnal Transcriptome from Chlamydomonas reinhardtii Reveals Continuous Cell and Metabolic Differentiation. *Plant Cell* **27,** 2743–2769 (2015).
39. Strenkert, D. *et al.* Multiomics resolution of molecular events during a day in the life of Chlamydomonas. *Proc. Natl. Acad. Sci. U.S.A.* **116,** 2374–2383 (2019).
40. Fett, J. P. & Coleman, J. R. Regulation of Periplasmic Carbonic Anhydrase Expression in Chlamydomonas reinhardtii by Acetate and pH. *Plant Physiol.* **106,** 103–108 (1994).

41. Pérez-Rodríguez, P. *et al.* PlnTFDB: Updated content and new features of the plant transcription factor database. *Nucl. Acids Res.* **38,** (2009).
42. Jin, J. *et al.* PlantTFDB 4.0: Toward a central hub for transcription factors and regulatory interactions in plants. *Nucl. Acids Res.* **45,** D1040–D1045 (2017).
43. Fischer, B. B. *et al.* SINGLET OXYGEN RESISTANT 1 links reactive electrophile signaling to singlet oxygen acclimation in Chlamydomonas reinhardtii. *Proc. Natl. Acad. Sci. U.S.A.* **109,** E1302–11 (2012).
44. Margolin, A. A. *et al.* ARACNE: An algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics* **7,** 1–15 (2006).
45. Barzel, B. & Barabási, A.-L. Network link prediction by global silencing of indirect correlations. *Nature Biotechnology* **31,** 720–725 (2013).
46. Omranian, N., Eloundou-Mbebi, J. M. O., Mueller-Roeber, B. & Nikoloski, Z. Gene regulatory network inference using fused LASSO on multiple data sets. *Sci Rep* **6,** 20533 (2016).
47. de Borda, J.-C. Mémoire sur les élections au scrutin. *Histoire de lAcademie Royale des Sciences* (1781).
48. Li, X. *et al.* A genome-wide algal mutant library and functional screen identifies genes required for eukaryotic photosynthesis. *Nat. Genet.* **51,** 627–635 (2019).
49. UniProt Consortium. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.* **49,** D480–D489 (2021).
50. Mi, H. *et al.* PANTHER version 16: A revised family classification, tree-based classification tool, enhancer regions and extensive API. *Nucl. Acids Res.* **49,** D394–D403 (2021).
51. Thiriet-Rupert, S. *et al.* Transcription factors in microalgae: genome-wide prediction and comparative analysis. *BMC Genomics* **17,** 282 (2016).
52. Bushnell, B. BBMap: A Fast, Accurate, Splice-Aware Aligner. *Lawrence Berkeley National Laboratory. LBNL Report LBNL-E* (2014).
53. Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. Voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biology* **15,** 1–17 (2014).
54. Anders, S., Pyl, P. T. & Huber, W. HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics* **31,** 166–169 (2015).
55. Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology* **11,** 1–9 (2010).
56. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26,** 139–140 (2009).
57. Huynh-Thu, V. A., Irrthum, A., Wehenkel, L. & Geurts, P. Inferring regulatory networks from expression data using tree-based methods. *PLoS ONE* **5,** e12776 (2010).
58. Meyer, P. E., Lafitte, F. & Bontempi, G. Minet: A r/bioconductor package for inferring large transcriptional networks using mutual information. *BMC Bioinformatics* **9,** 1–10 (2008).
59. Olsen, C., Meyer, P. E. & Bontempi, G. On the impact of entropy estimation on transcriptional regulatory network inference based on mutual information. *Eurasip Journal on Bioinformatics and Systems Biology* **2009,** (2009).
60. Faith, J. J. *et al.* Large-scale mapping and validation of Escherichia coli transcriptional regulation from a compendium of expression profiles. *PLoS Biol.* **5,** 0054–0066 (2007).
61. Feizi, S., Marbach, D., Médard, M. & Kellis, M. Network deconvolution as a general method to distinguish direct dependencies in networks. *Nature Biotechnology* **31,** 726–733 (2013).
62. Ritchie, M. E. *et al.* Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucl. Acids Res.* **43,** e47 (2015).
63. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15,** 1–21 (2014).

619 64. North, B. V., Curtis, D. & Sham, P. C. A note on the calculation of empirical P values from Monte Carlo
620 procedures [1]. *American Journal of Human Genetics* **71,** 439–441 (2002).
621 65. Gorman, D. S. & Levine, R. P. Cytochrome f and plastocyanin: their sequence in the photosynthetic
622 electron transport chain of *Chlamydomonas reinhardii*. *Proc. Natl. Acad. Sci. U.S.A.* **54,** 1665–1669
623 (1965).
624 66. Sueoka, N. Mitotic replication of deoxyribonucleic acid in *Chlamydomonas reinhardii*. *Proc. Natl.*
625 *Acad. Sci. U.S.A.* **46,** 83–91 (1960).
626 67. Greiner, A. *et al.* Targeting of Photoreceptor Genes in Chlamydomonas reinhardtii via Zinc-Finger
627 Nucleases and CRISPR/Cas9. *Plant Cell* **29,** 2498–2518 (2017).
628 68. Kaye, Y. *et al.* The mitochondrial alternative oxidase from Chlamydomonas reinhardtii enables
629 survival in high light. *J. Biol. Chem.* **294,** 1380–1395 (2019).
630 69. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat.*
631 *Methods* **6,** 343–345 (2009).
632 70. Mackinder, L. C. M. *et al.* A repeat protein links Rubisco to form the eukaryotic carbon-concentrating
633 organelle. *Proc. Natl. Acad. Sci. U.S.A.* **113,** 5958–5963 (2016).
634 71. Schloss, J. A. A Chlamydomonas gene encodes a G protein beta subunit-like polypeptide. *Molec. Gen.*
635 *Genet.* **221,** 443–452 (1990).
636 72. Schlesselman, J. J. Data Transformation in Two-Way Analysis of Variance. *Journal of the American*
637 *Statistical Association* (2012). doi:10.1080/01621459.1973.10482435
638 73. Genty, B., Briantais, J.-M. & Baker, N. R. The relationship between the quantum yield of
639 photosynthetic electron transport and quenching of chlorophyll fluorescence. *Biochimica et*
640 *Biophysica Acta (BBA)* **990,** 87–92 (1989).
641

Acknowledgments

We are grateful to Prof. Hideya Fukuzawa for sending us the *lcr1* and *lcr1-C* strains and to Prof. Peter Jahns for the antibody against PSBS

Funding:

The authors would like to thank the following agencies for funding: The Human Frontiers Science Program through the funding of the project RGP0046/2018 (DP, ZN); the French National Research Agency in the framework of the Young Investigators program ANR-18-CE20-0006 through the funding of the project MetaboLight (DP); the French National Research Agency in the framework of the Investissements d'Avenir program ANR-15-IDEX-02, through the funding of the "Origin of Life" project of the Univ. Grenoble-Alpes (DP, YY); the French National Research Agency through the funding of the Grenoble Alliance for Integrated Structural & Cell Biology GRAL project ANR-17-EURE-0003 (DP, MAR-S), the Prestige Marie-Curie co-financing grant PRESTIGE-2017-1-0028 (MAR-S); the International Max Planck Research School 'Primary Metabolism and Plant Growth' at the Max Planck Institute of Molecular Plant Physiology (MA, ZN).

Competing interests: Authors declare that they have no competing interests.

663 **Figures**

664

665 **Fig. 1: Characterization of the consensus GRN inferred by employing a compendium of RNA-seq**
666 **data from diverse light and culture conditions**. **a.** Expression levels of representative CCM and
667 qE related genes are plotted over all samples used for network inference (z-scaled log values
668 are depicted). The column annotation gives information on the culture conditions. The values
669 of the rows of the heatmap correspond to the z-scores of representative CCM and qE genes
670 plotted over all samples (columns) used in the network inference. **b.** The heatmap rows
671 correspond to experimentally validated or falsified (neg) gene regulatory interactions involved
672 in qE and CCM, curated from literature. The heatmap indicates ranking of these interactions by
673 different approaches and the consensus network. Edges are considered highly ranked (depicted
674 in blue) if they are above the 10% network density threshold. Edges ranked below this threshold
675 are depicted in red. Edges that were not included in the given network are marked in grey.
676 ARACNE and Silencing columns were only plotted for comparison and were not used in building
677 the consensus GRN (see Methods).

678 **Fig. 2: Consensus GRN for light and acetate responses pinpoints LCR1 as regulator of qE-related**
679 **genes. a.** Dot plot of the relative regulatory strength of the top 10 regulators of qE-related
680 genes in the consensus GRN (see Methods). TFs are marked in green if qE transcript levels were
681 affected in the respective knock-out strain and this effect was reversed by complementation
682 with the missing gene. TFs for which no effect was observed are marked in red. TFs for which
683 no mutant lines were available are plotted in grey. **b.** WT, *lcr1* and *lcr1-C* cells were acclimated
684 for 16h in LL (15 µmol photons m^{-2} s^{-1}). After sampling for the LL conditions, light intensity was
685 increased to 300 µmol photons m^{-2} s^{-1} (HL); samples were taken 1 h (RNA) or 4 h (protein and
686 photosynthetic measurements) after exposure to HL. Shown are relative expression levels of
687 qE-related genes at the indicated conditions normalized to WT LL (*n* = 3 biological samples,
688 mean ± sd). **c**. Immunoblot analyses of LHCSR1, LHCSR3, PSBS and ATPB (loading control) of one
689 of the three biological replicates, under the indicated conditions. **d.** Quantification of
690 immunoblot data of all replicates in panel **c** after normalization to ATPB. Shown are the HL
691 treated samples; WT protein levels were set as 1. **e.** NPQ and calculated qE (as an inset) 4h after
692 exposure to HL (*n* = 3 biological samples, mean ± s.d). The p-values for the comparisons are
693 based on ANOVA Dunnett's multiple comparisons test and as indicated in the graphs (*, P <
694 0.005, **, P < 0.01, ***, P < 0.001, ****, P < 0.0001). Statistical analyses for panel **b** and **d** were
695 applied on log10- transformed values.

696 **Fig. 3: A PHOT-specific GRN pinpoints QER7 as a suppressor of the expression of qE-related genes.**
697 **a.** Dot plot of the relative regulatory strength of the top 10 regulators of qE-related genes in
698 the PHOT-specific GRN (see **Methods**). TFs are marked in green if qE transcript levels were
699 affected in the respective knock-out strain and this effect was reversed by complementation
700 with the knocked-out gene, in yellow, if the effect was not reversed by complementation, and
701 in red, if no mutant effect was observed in the mutant. TFs for which no mutant lines were
702 available are plotted in grey. **b**. WT, *qer7* and *qer7-C* cells were synchronized under 12h light
703 (15 µmol m^{-2} s^{-1})/12h dark cycles. After sampling for the dark conditions (end of the dark phase),
704 cells were exposed to 300 µmol photons m^{-2} s^{-1} (HL); samples were taken 1 h (RNA) or 4 h
705 (protein and photosynthetic measurements) after exposure to HL. Shown are relative
706 expression levels of qE-related genes at the indicated conditions normalized to WT LL (*n* = 3

707 biological samples, mean ± sd). **c**. Immunoblot analyses of LHCSR1, LHCSR3 and ATPB (loading
708 control) of one of the three biological replicate set of samples, under the indicated conditions.
709 **d.** Quantification of immunoblot data of all replicates in panel **c** after normalization to ATPB.
710 WT protein levels at HL were set as 1. **e.** NPQ and qE, measured 4h after exposure to HL (*n* = 3
711 biological samples, mean ± sd). The p-values for the comparisons are based on ANOVA
712 Dunnett's multiple comparisons test and are indicated in the graphs (*, P < 0.005, **, P < 0.01,
713 ***, P < 0.001, ****, P < 0.0001). Statistical analyses for panel **b** and **d** were applied on log10-
714 transformed values.

715 **Fig:4: QER7 suppresses transcription of CCM genes and depends on PHOT.** WT, *qer7* and *qer7-C* cells
716 were synchronized under 12h light (15 µmol m^{-2} s^{-1})/12h dark cycles. After sampling at the end
717 of the dark phase, cells were exposed to 300 µmol photons m^{-2} s^{-1} (HL) and samples were taken
718 1 h after HL exposure. **a.** Relative expression levels of CCM genes at the indicated conditions
719 normalized to WT LL (*n* = 3 biological samples, mean ± sd). **b and c.** Relative expression levels
720 of *QER7* (**b**) and *LCR1* (**c**) in synchronized WT, *phot* and *phot-C* (phot cells complemented with
721 the *PHOT* gene) cells. The p-values for the comparisons are based on ANOVA Dunnett's multiple
722 comparisons test on log10- transformed values and are indicated in the graphs (*, P < 0.005,
723 **, P < 0.01, ***, P < 0.001, ****, P < 0.0001).

724 **Fig. 5: Consensus and PHOT-specific GRNs indicate extensive co-regulation of CCM and**
725 **photoprotective genes.** Top: Venn diagram depicting the overlap of the top 10 predicted TFs
726 of the curate CCM and photoprotective (PhPr) genes based on **a.** the consensus or **c.** PHOT-
727 specific GRN. Bottom: Network representation of the top 10 TF sets of **b.** the consensus
728 network or **d.** the phot GRN (center nodes, same color code as in panel **a** and the target genes
729 used for prediction (left and right columns of nodes, photoprotection genes are shown in green,
730 qE-related genes in blue, and CCM genes in red. Larger TF nodes indicate higher regulatory
731 strength on target genes. The plotted edge width is proportional to the strength of the
732 regulatory interaction.

733

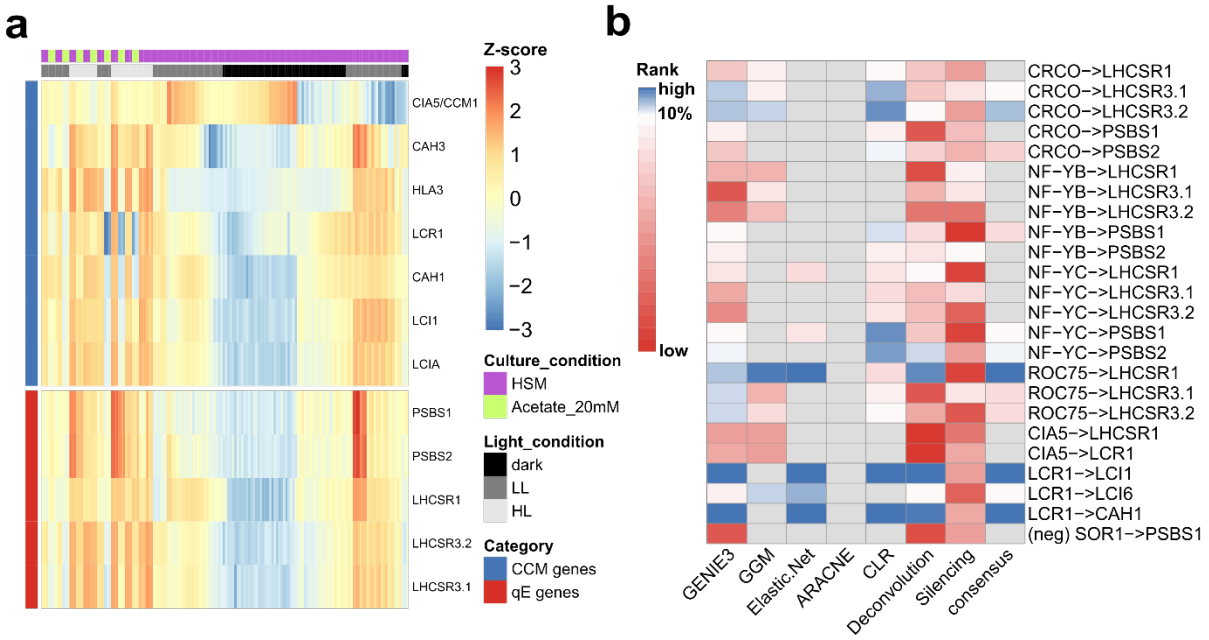


Fig. 1: Characterization of the consensus GRN inferred by employing a compendium of RNA-seq data from diverse light and culture conditions. **a.** Expression levels of representative CCM and qE related genes are plotted over all samples used for network inference (z-scaled log values are depicted). The column annotation gives information on the culture conditions. The values of the rows of the heatmap correspond to the z-scores of representative CCM and qE genes plotted over all samples (columns) used in the network inference. **b.** The heatmap rows correspond to experimentally validated or falsified (neg) gene regulatory interactions involved in qE and CCM, curated from literature. The heatmap indicates ranking of these interactions by different approaches and the consensus network. Edges are considered highly ranked (depicted in blue) if they are above the 10% network density threshold. Edges ranked below this threshold are depicted in red. Edges that were not included in the given network are marked in grey. ARACNE and Silencing columns were only plotted for comparison and were not used in building the consensus GRN (see Methods).

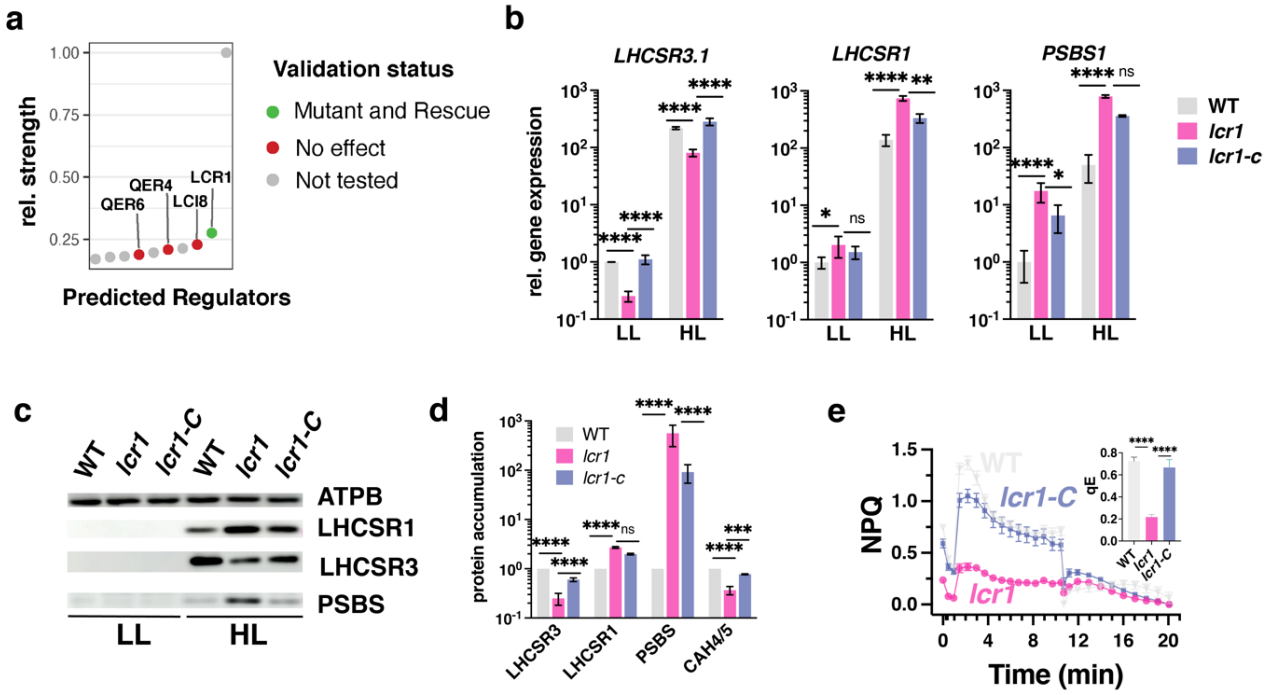


Fig. 2: Consensus GRN for light and acetate responses pinpoints LCR1 as regulator of qE-related genes. a. Dot plot of the relative regulatory strength of the top 10 regulators of qE-related genes in the consensus GRN (see Methods). TFs are marked in green if qE transcript levels were affected in the respective knock-out strain and this effect was reversed by complementation with the missing gene. TFs for which no effect was observed are marked in red. TFs for which no mutant lines were available are plotted in grey. **b.** WT, *lcr1* and *lcr1-C* cells were acclimated for 16h in LL (15 µmol photons m^{-2} s^{-1}). After sampling for the LL conditions, light intensity was increased to 300 µmol photons m^{-2} s^{-1} (HL); samples were taken 1 h (RNA) or 4 h (protein and photosynthetic measurements) after exposure to HL. Shown are relative expression levels of qE-related genes at the indicated conditions normalized to WT LL (n = 3 biological samples, mean ± sd). **c**. Immunoblot analyses of LHCSR1, LHCSR3, PSBS and ATPB (loading control) of one of the three biological replicates, under the indicated conditions. **d.** Quantification of immunoblot data of all replicates in panel **c** after normalization to ATPB. Shown are the HL treated samples; WT protein levels were set as 1. **e.** NPQ and calculated qE (as an inset) 4h after exposure to HL (n = 3 biological samples, mean ± s.d). The p-values for the comparisons are based on ANOVA Dunnett's multiple comparisons test and as indicated in the graphs (*, P < 0.005, **, P < 0.01, ***, P < 0.001, ****, P < 0.0001). Statistical analyses for panel **b** and **d** were applied on log10-transformed values.

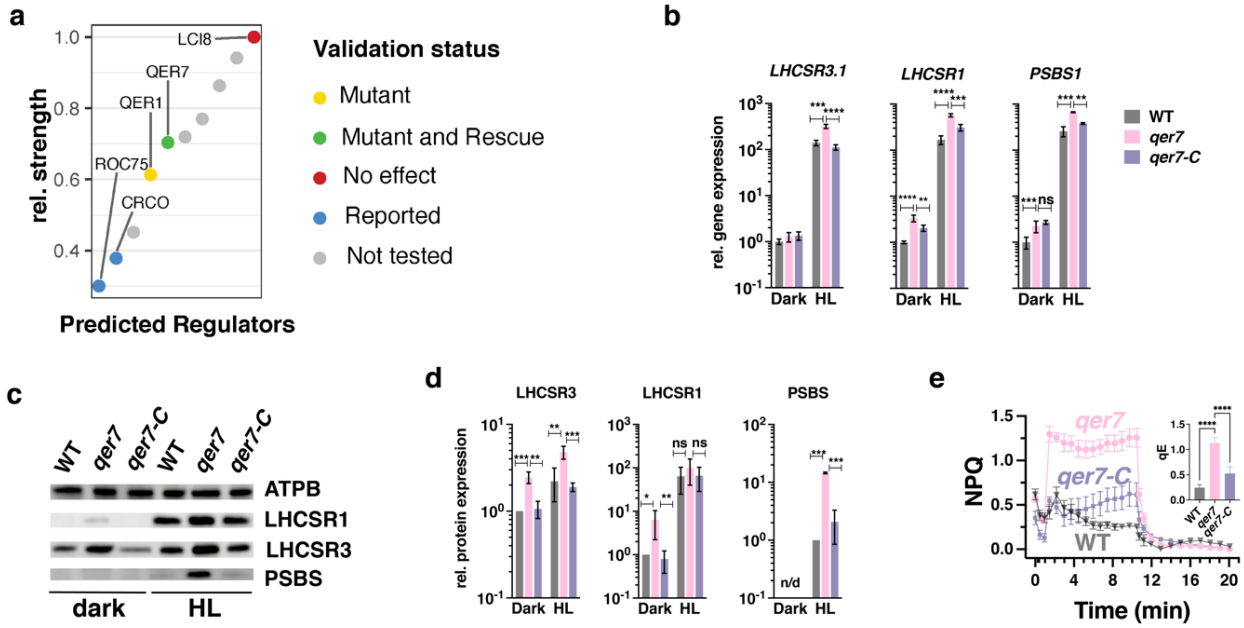
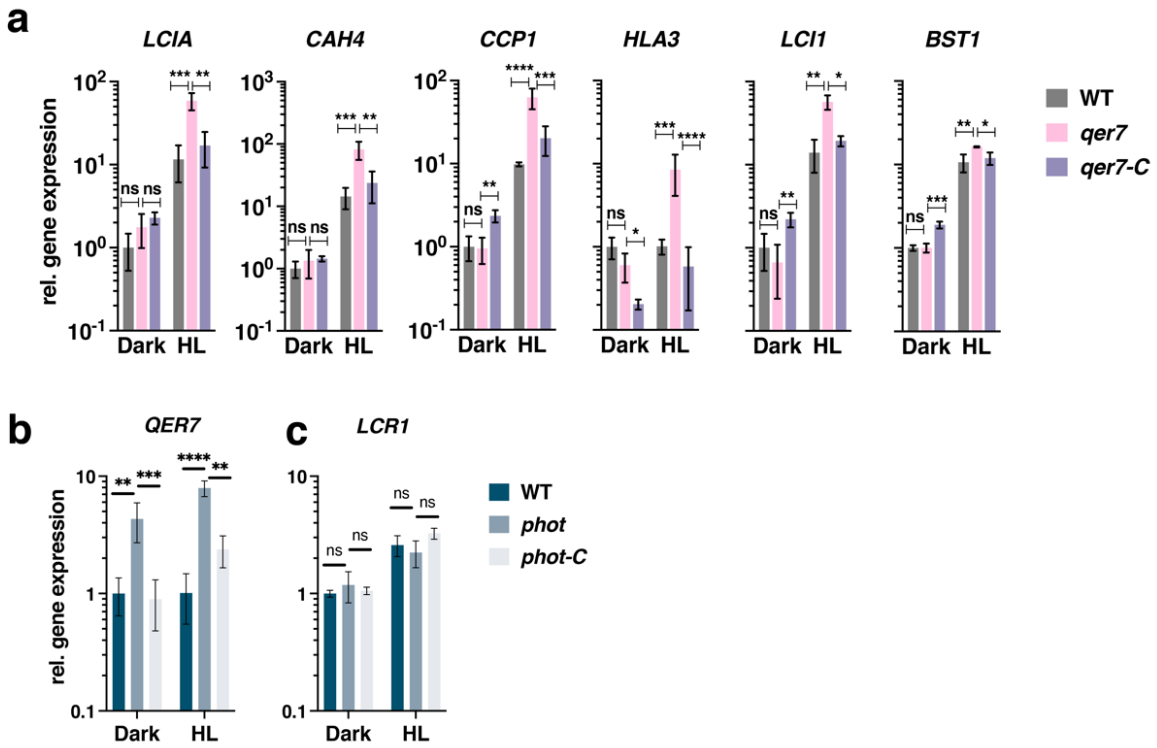


Fig. 3: A PHOT-specific GRN pinpoints QER7 as a suppressor of the expression of qE-related genes. a. Dot plot of the relative regulatory strength of the top 10 regulators of qE-related genes in the PHOT-specific GRN (see **Methods**). TFs are marked in green if qE transcript levels were affected in the respective knock-out strain and this effect was reversed by complementation with the knocked-out gene, in yellow, if the effect was not reversed by complementation, and in red, if no mutant effect was observed in the mutant. TFs for which no mutant lines were available are plotted in grey. **b**. WT, *qer7* and *qer7-C* cells were synchronized under 12h light (15 µmol m^{-2} s^{-1})/12h dark cycles. After sampling for the dark conditions (end of the dark phase), cells were exposed to 300 µmol photons m^{-2} s^{-1} (HL); samples were taken 1 h (RNA) or 4 h (protein and photosynthetic measurements) after exposure to HL. Shown are relative expression levels of qE-related genes at the indicated conditions normalized to WT LL (*n* = 3 biological samples, mean ± sd). **c**. Immunoblot analyses of LHCSR1, LHCSR3 and ATPB (loading control) of one of the three biological replicate set of samples, under the indicated conditions. **d.** Quantification of immunoblot data of all replicates in panel **c** after normalization to ATPB. WT protein levels at HL were set as 1. **e.** NPQ and qE, measured 4h after exposure to HL (*n* = 3 biological samples, mean ± sd). The p-values for the comparisons are based on ANOVA Dunnett's multiple comparisons test and are indicated in the graphs (*, P < 0.005, **, P < 0.01, ***, P < 0.001, ****, P < 0.0001). Statistical analyses for panel **b** and **d** were applied on log10- transformed values.



787

788 **Fig:4: QER7 suppresses transcription of CCM genes and depends on PHOT.** WT, *qer7* and *qer7-C* cells were
789 synchronized under 12h light (15 µmol m^{-2} s^{-1})/12h dark cycles. After sampling at the end of the dark phase,
790 cells were exposed to 300 µmol photons m^{-2} s^{-1} (HL) and samples were taken 1 h after HL exposure. **a.** Relative
791 expression levels of CCM genes at the indicated conditions normalized to WT LL (*n* = 3 biological samples,
792 mean ± sd). **b and c.** Relative expression levels of *QER7* (**b**) and *LCR1* (**c**) in synchronized WT, *phot* and *phot-C*
793 (phot cells complemented with the *PHOT* gene) cells. The p-values for the comparisons are based on ANOVA
794 Dunnett's multiple comparisons test on log10- transformed values and are indicated in the graphs (*, P <
795 0.005, **, P < 0.01, ***, P < 0.001, ****, P < 0.0001).

796

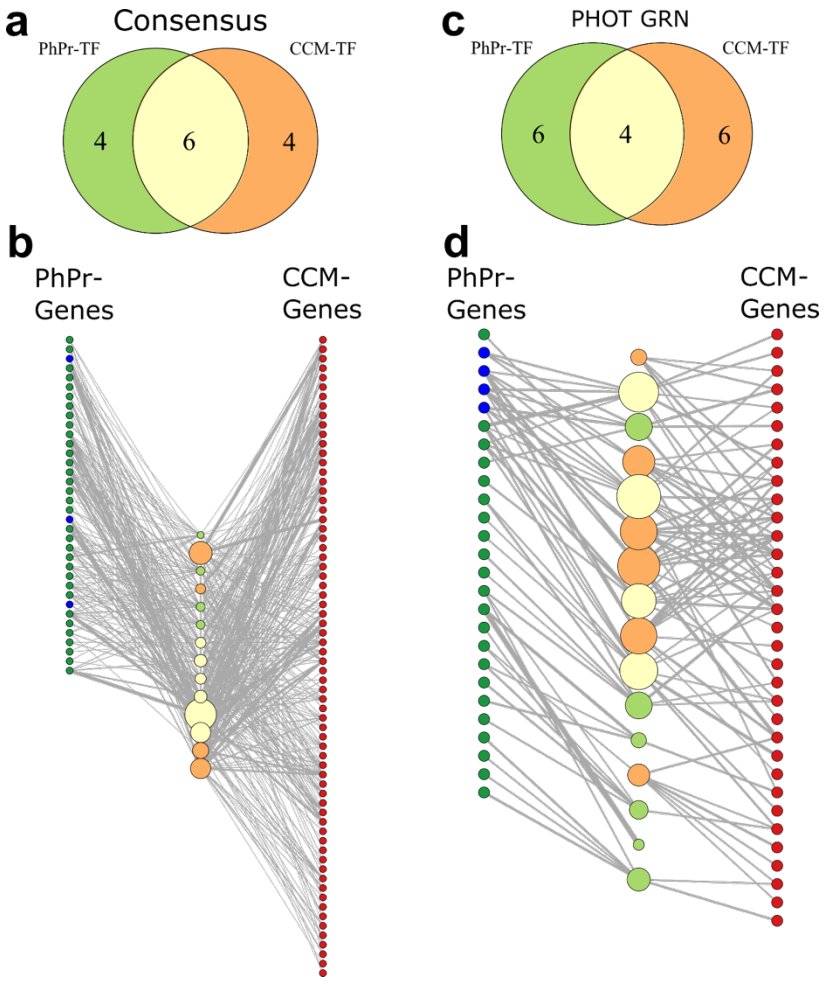


Fig. 5: Consensus and PHOT-specific GRNs indicate extensive co-regulation of CCM and photoprotective genes. Top: Venn diagram depicting the overlap of the top 10 predicted TFs of the curate CCM and photoprotective (PhPr) genes based on **a.** the consensus or **c.** PHOT-specific GRN. Bottom: Network representation of the top 10 TF sets of **b.** the consensus network or **d.** the phot GRN (center nodes, same color code as in panel **a** and the target genes used for prediction (left and right columns of nodes, photoprotection genes are shown in green, qE-related genes in blue, and CCM genes in red. Larger TF nodes indicate higher regulatory strength on target genes. The plotted edge width is proportional to the strength of the regulatory interaction.

Supplementary Materials for

Widening the landscape of transcriptional regulation of algal photoprotection

Authors

Marius Arend[1,2†], Yizhong Yuan[3†], M. Águila Ruiz-Sola[3‡], Nooshin Omranian[1,2,4], Zoran Nikoloski[1,2,4*], Dimitris Petroutsos[3 *]

[†] these authors contributed equally to this work

[*]**Corresponding Authors**: nikoloski@mpimp-golm.mpg.de (Z.N.), dimitris.petroutsos@cea.fr (D.P.)

This PDF file includes:

Supplementary Note

Extended Data Figs. 1 to 10

Other Supplementary Materials for this manuscript include the following:

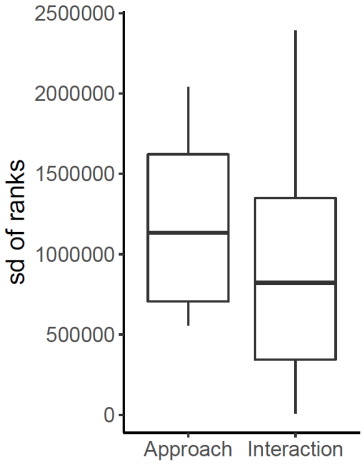
Tables S1 to S10 (to be downloaded as excel files)

30 **Supplementary Note**

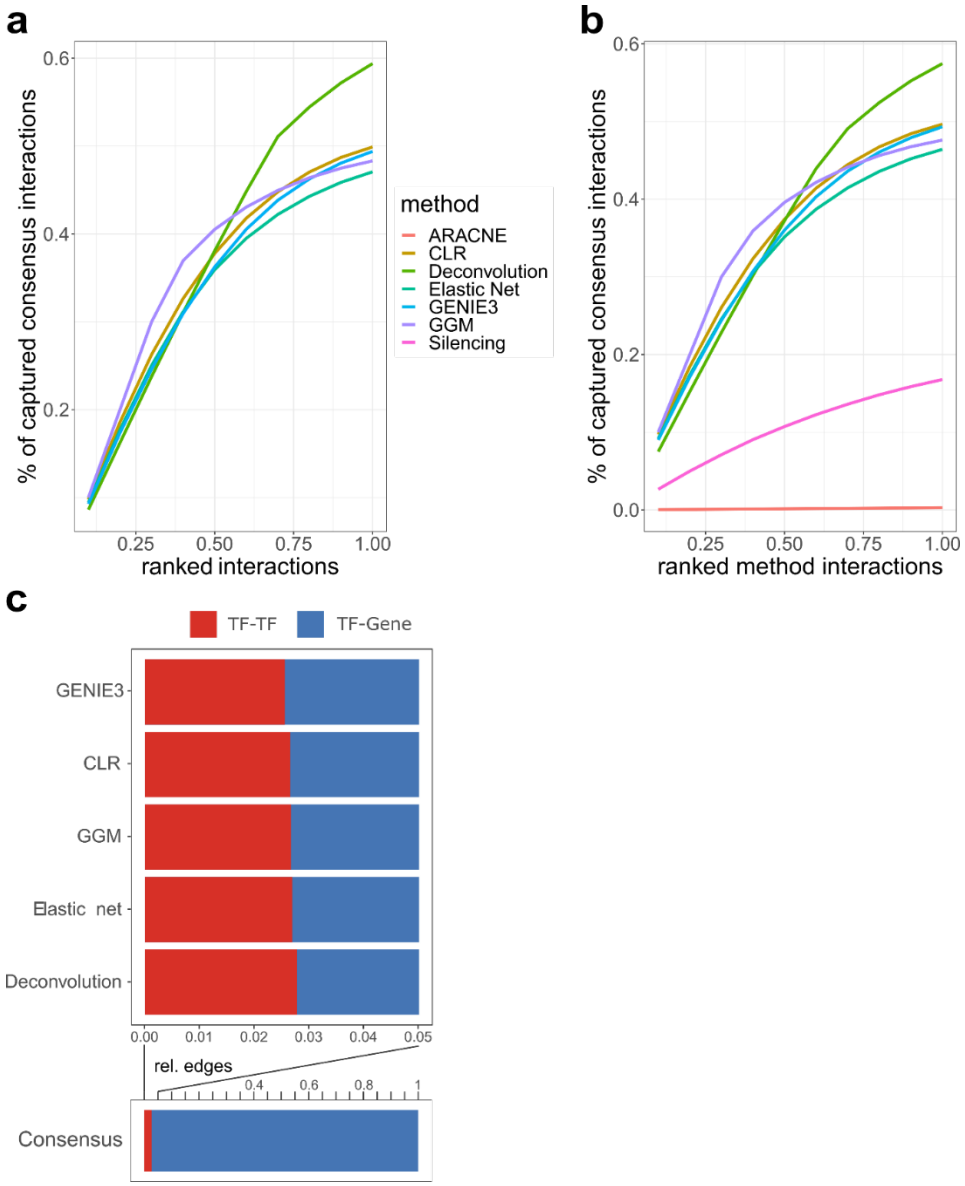
31 Simulation studies[30] indicated that the top 10% of the edges in the consensus network are enriched in positive
32 interactions if the underlying approaches perform better than guessing. Thus, a higher overlap with the
33 consensus networks is expected to result in improved predictions. We found that the interactions from the
34 network deconvolution approach exhibited the largest overlap with the consensus network, while GGM
35 assigns the highest rank to consensus interactions (**Extended Data Fig. 2a**). Repeating this analysis by
36 including the two worse performing approaches (i.e. ARACNE and Silencing) resulted in no qualitative changes
37 in the overlaps (**Extended Data Fig. 2b**) and only 7.47% of difference to the original consensus network,
38 demonstrating the robustness of the inferred interactions. Considering the 10% network density threshold,
39 we also inspected the proportion of TF-TF interactions in the different networks and their consensus; we
40 found that it ranges from 2.55% in the GENIE3 network to 2.77% in the network deconvolution approach
41 (**Extended Data Fig. 2c**). Since the fully connected network containing all possible TF-TF and TF-target
42 interactions has a relative TF-TF content of 2.52%, these findings suggest an enrichment of TF-TF interactions
43 in GRNs inferred by all five approaches considered in the consensus.

44

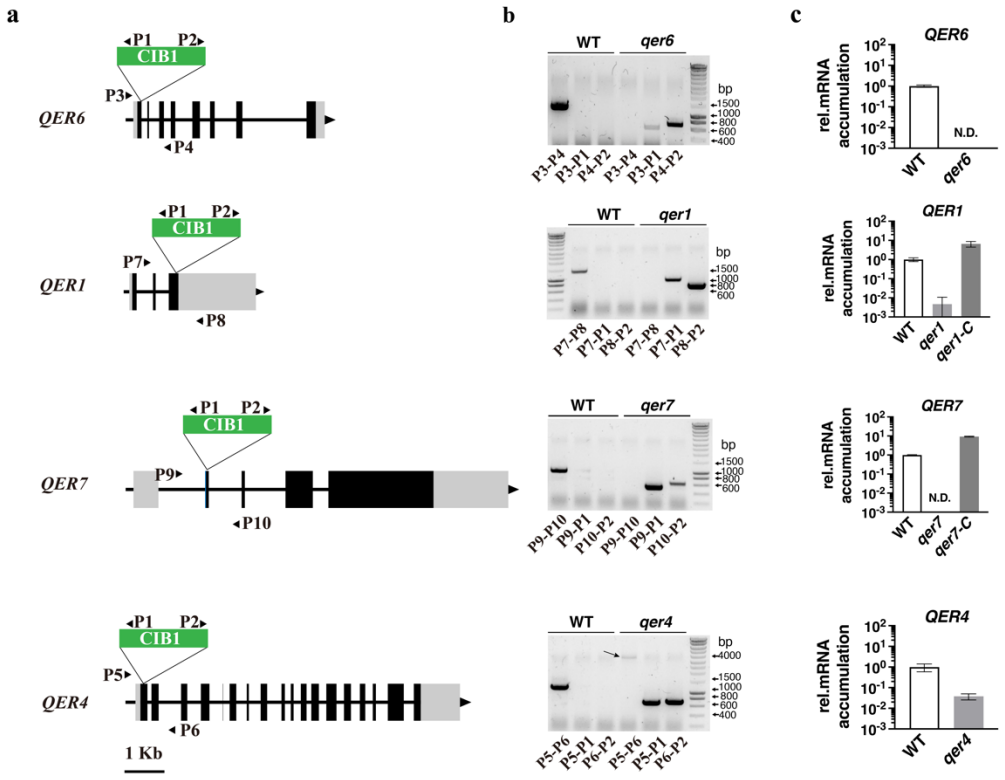
45



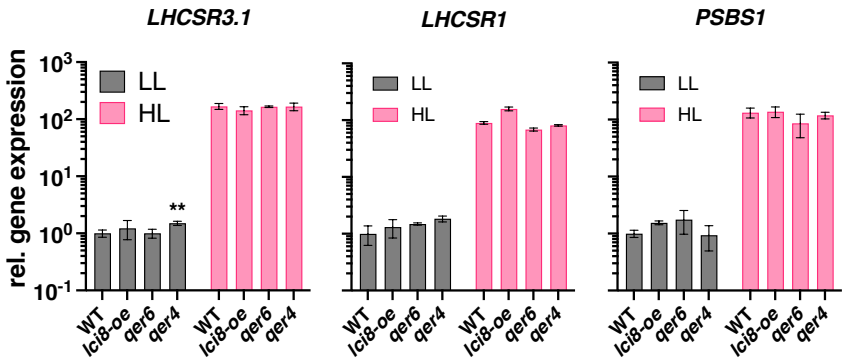
Extended Data Fig. 1: Variation in ranks of known interactions across approaches and interactions from different inference approaches. Boxplots of the standard deviation of ranks calculated for the curated interactions plotted in **Fig. 1b** over all ranks assigned by one approach or over all ranks assigned to one interaction.



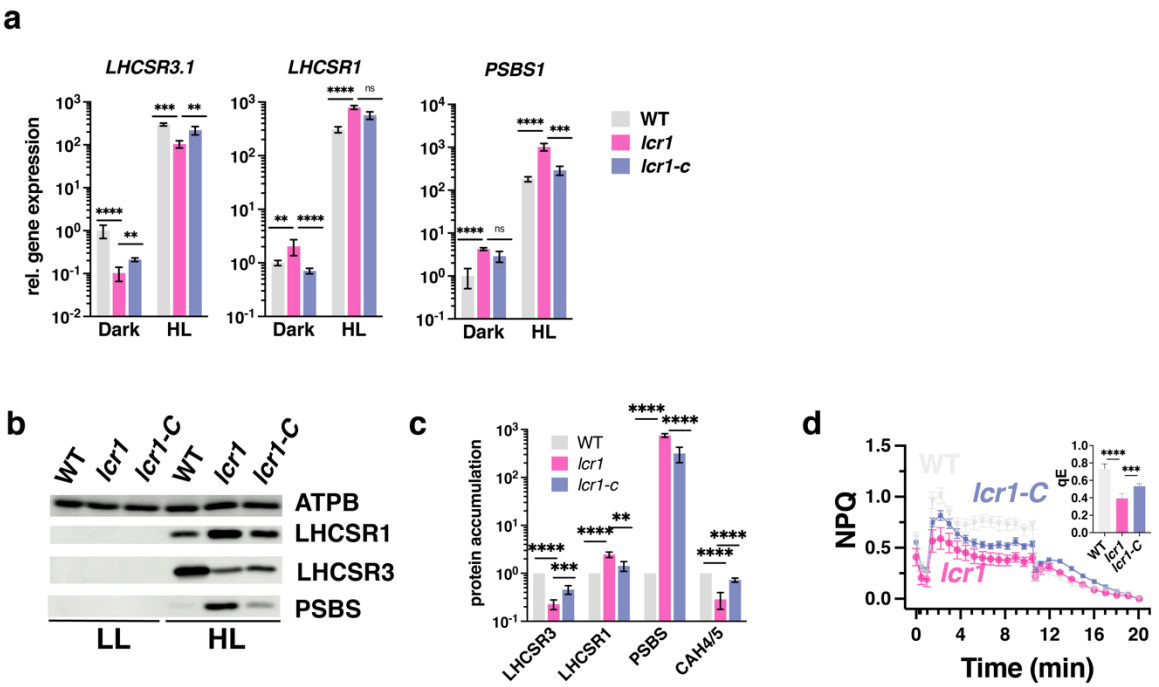
Extended Data Fig. 2: Overlap of consensus network with individual inference methods and proportion of TF-TF interactions in the obtained GRNs. a. The graph shows the overlap between the edges of the consensus network and the ranked interactions of the individual approaches normalized to the total number of edges in the consensus network used for regulator prediction. **b.** Same as in a but plotting the overlap with a consensus of all used approaches (including ARACNE and Silencing). **c**. Bar plots provide the proportion of TF-TF interactions in comparison to TF-target gene interactions contained in the GRNs inferred by the different approaches and the consensus GRN resulting from their integration. All depicted analyses considered only the interactions within in the 10% network density threshold (see **Methods**).



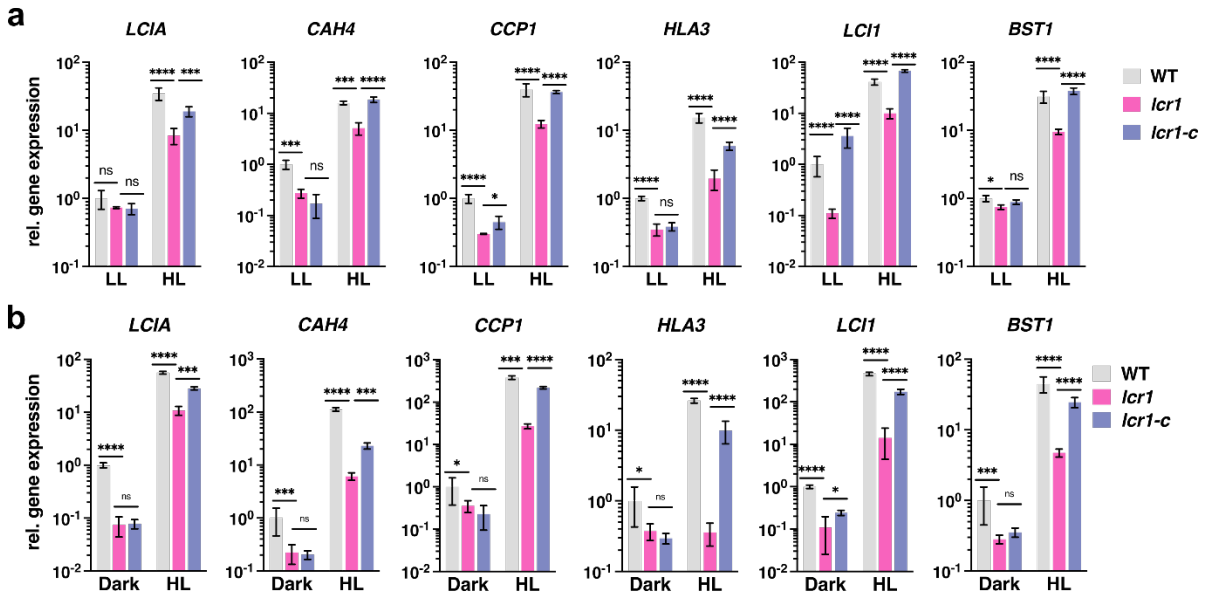
Extended Data Fig. 3: Genotyping of CLiP mutants affected in the predicted regulators of qE-related genes. **a**. Insertion map of the CIB1 cassette in the different genes. Exons are shown in black, introns as interconnecting lines, 5'UTR and 3'UTR in light gray, and primers in arrows. The insertion site of the CIB1 cassette is indicated by the triangle; **b**. PCR-validation of the insertion site in the different CLiP mutants using genomic DNA. To confirm the CIB1 insertion site, gene-specific primers were used that anneal upstream and downstream of the predicted insertion site of the cassette (primer pairs P3-P4, P7-P8, P9-P10 and P5-P6 for *qer6*, *qer1*, *qer7* and *qer4* respectively*;* **Supplementary Table 9**). Pairs of primers used are indicated at the bottom of the agarose gels used to separate the PCR products. Note that the PCR product of P5-P6 is indicated by an arrow. **c.** Relative expression levels of predicted qE regulator genes in the different CLiP mutants after exposure to HL (300 µmol photons m^{-2} s^{-1}) for 1h.



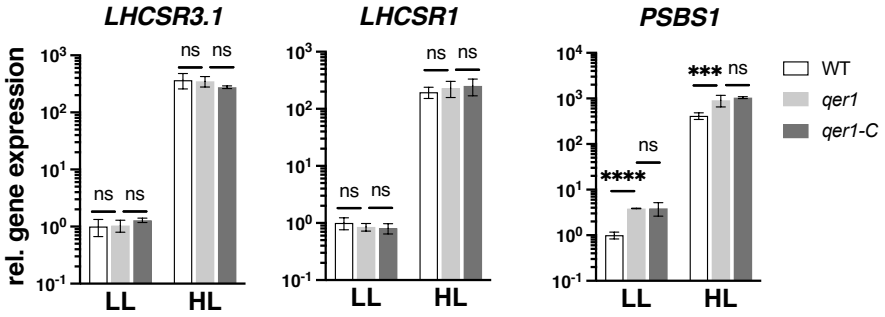
Extended Data Fig. 4: qE gene and protein expression in CLiP mutants bearing mutations in predicted qE regulator genes. WT, *lci8-oe, qer6* and *qer4* cells were acclimated for 16h in LL (15 µmol photons m^{-2} s^{-1}). After sampling for the LL conditions, light intensity was increased to 300 µmol photons m^{-2} s^{-1} (HL); samples were taken 1 h after exposure to HL. Shown are relative expression levels of of *LHCSR3.1*, *LHCSR1* and *PSBS1* at the indicated conditions normalized to WT LL (*n* = 3 biological samples, mean ± sd). The p-values for the comparisons between the mutants and the WT are based on ANOVA Dunnett's multiple comparisons test on log10- transformed values and are indicated in the graphs (**, $P < 0.01$).



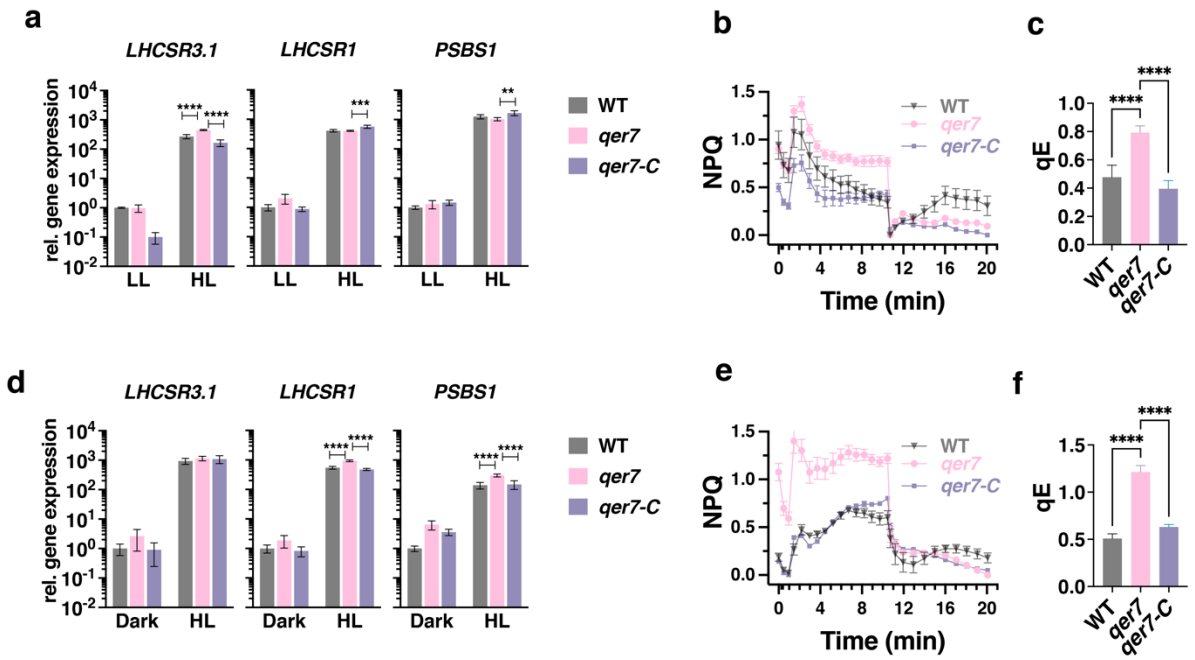
Extended Data Fig. 5: LCR1 is required for proper expression of qE-related genes during transitions from dark-to-light. WT, *lcr1* and *lcr1-C* cells were acclimated for 16h in darkness. After sampling for the dark conditions, light intensity was increased to 300 µmol photons m^{-2} s^{-1} (HL); samples were taken 1 h (RNA) or 4 h (protein and photosynthetic measurements) after exposure to HL. **a**. Relative expression levels of qE genes at the indicated conditions normalized to WT LL (*n* = 3 biological samples, mean ± sd). **b**. Immunoblot analyses of LHCSR1, LHCSR3, PSBS and ATPB (loading control) of one out of the three biological replicate samples, under the indicated conditions. **c.** Summary graph of immunoblots of all replicate samples of Extended Data Fig. 3b after normalization to ATPB. Shown are the HL treated samples; WT protein levels were set as 1. **d.** NPQ and calculated qE 4h after exposure to HL (*n* = 3 biological samples, mean ± s.d). The p-values for the comparisons are based on ANOVA Dunnett's multiple comparisons test and are indicated in the graphs (*, P < 0.005, **, P < 0.01, ***, P < 0.001, ****, P < 0.0001). Statistical analyses for panel **a** and **c** were applied on log10- transformed values.



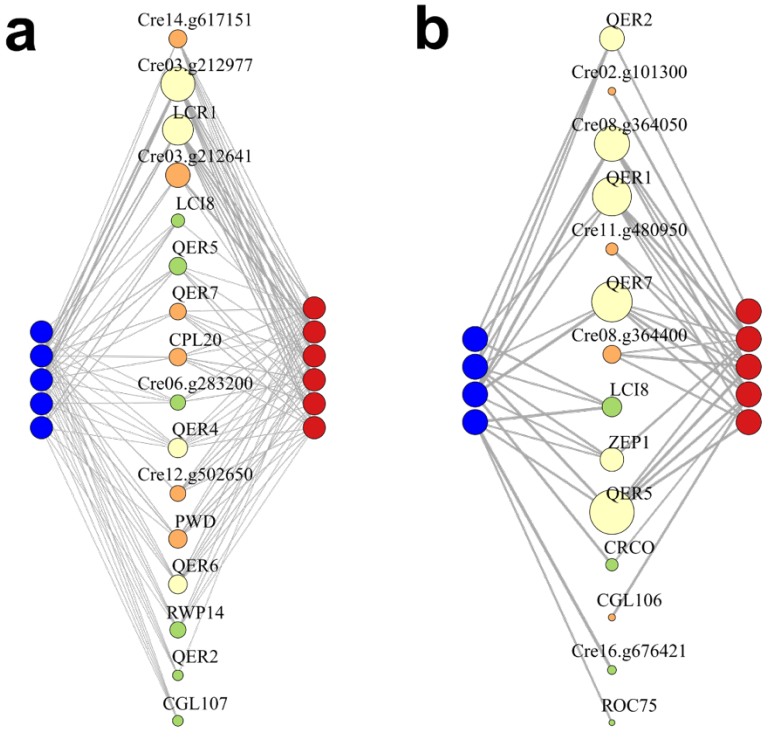
Extended Data Fig. 6: LCR1 activates the transcription of CCM genes in LL and dark-acclimated cells. a. WT, *lcr1* and *lcr1-C* cells were acclimated for 16h in LL (15 µmol photons m^{-2} s^{-1}). After sampling under LL conditions, light intensity was increased to 300 µmol photons m^{-2} s^{-1} (HL) and samples were taken 1 h after exposure to HL. Relative expression of CCM genes at the indicated conditions normalized to WT under LL. **b**. Relative expression of CCM genes acclimated in 16h of darkness (*n* = 3 biological samples, mean ± sd). The p-values for the comparisons are based on ANOVA Dunnett's multiple comparisons test on log10- transformed values and are indicated in the graphs (*, P < 0.005, **, P < 0.01, ***, P < 0.001, ****, P < 0.0001).



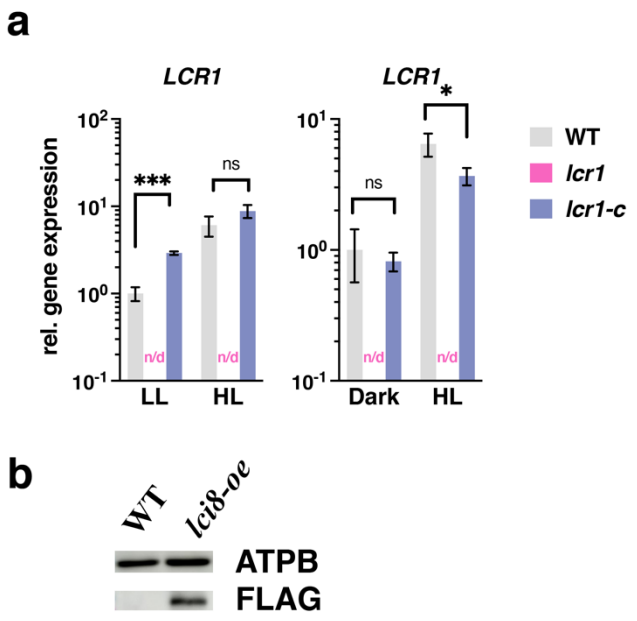
Extended Data Fig.7: Complementation of *qer1* with *QER1* (*qer1-C* strain) fails to rescue the *LHCSR1* and *PSBS* phenotypes. WT, *qer1* and *qer1-C* cells were acclimated for 16h in LL (15 µmol photons m^{-2} s^{-1}). After sampling for the dark conditions, light intensity was increased to 300 µmol photons m^{-2} s^{-1} (HL); RNA samples were taken 1 h after exposure to HL. Shown are relative expression levels of qE genes at the indicated conditions normalized to WT LL (*n* = 3 biological samples, mean ± sd). The p-values for the comparisons are based on ANOVA Dunnett's multiple comparisons test on log10- transformed values and are indicated in the graphs (***, P < 0.001, ****, P < 0.0001).



Extended Data Fig.8: Relative expression of qE-related genes in asynchronous *qer7* cells. WT, *qer7* and *qer7-C* cells were acclimated for 16h in LL (15 µmol photons m^{-2} s^{-1}; **a-c**) or darkness (**d-f**). After sampling for the LL or dark conditions, light intensity was increased to 300 µmol photons m^{-2} s^{-1} (HL) and samples were taken 1 h (RNA) or 4 h (photosynthetic measurements) after exposure to HL. **a, d**. Relative expression of qE and CCM genes at the indicated conditions normalized to WT LL (**a**) or dark (**d**) respectively (*n* = 3 biological samples, mean ± sd). **b, e.** NPQ and **c, f.** calculated qE, 4h after exposure to HL (*n* = 3 biological samples, mean ± s.d. The p-values for the comparisons are based on ANOVA Dunnett's multiple comparisons test and are indicated in the graphs (**, P < 0.01, ***, P < 0.001, ****, P < 0.0001). Statistical analyses for panel **a** and **d** were applied on log_{10}- transformed values.



Extended Data Fig. 9: Reconstructed GRNs recapitulate experimentally found regulators of qE and CCM. Network representation of the top 10 regulators of CCM genes quantified via qPCR (**Fig. 4),** and qE genes as predicted from **a**. the concensus or **b**. PHOT-specific GRN. Regulators are represented by centeral nodes (same color code as in **Fig. 5**), target qE genes present in the respective network are shown in blue (*LHCSR1*, *LHCSR3.1*, *LHCSR3.2*, *PSBS1*, *PSBS2*) and target CCM genes are shown in red (*LCI1*, *LCIA*, *CAH4*, *CCP1*, *HLA3*, *BST1*). Larger TF nodes indicate higher regulatory strength on target genes. The plotted edge width is proportional to the strength of the regulatory interaction.



Extended Data Fig. 10: LCR1 mutation and LCI8 overexpression confirmation. **a.** Relative expression of *LCR1* in WT, *lcr1* and *lcr1-C* strains. Cells were acclimated for 16h in LL or darkness. After sampling for the LL or dark conditions, light intensity was increased to 300 µmol photons m^{-2} s^{-1} (HL), and samples for RNA purification were taken 1 h after exposure to HL (*n* = 3 biological samples, mean ± sd). *LCR1* gene expression was non-detectable (indicated as n/d in the graphs) in the *lcr1* mutant. **b.** Immunoblot analysis of LCI8-FLAG fused protein in WT and lci8-oe cells grown in TAP, LL. The p-values for the comparisons are based on unpaired t-tests on log10- transformed values and are indicated in the graphs (*, $P < 0.005$, ***, $P < 0.001$).

Supplementary Table Captions

Supplementary Table 1: RNAseq data sets used in this study; HSM: Sueoka's high salt medium (1); HSM – new trace = modified Sueoka's high salt medium with trace elements as described in (2); Rep= biological replicates

(1) Sueoka, N. Mitotic replication of deoxyribonucleic acid in Chlamydomonas reinhardii. Proc. Natl. Acad. Sci. U.S.A. 46, 83–91 (1960).

(2) Kropat, J.et al. A Revised Mineral Nutrient Supplement Increases Biomass and Growth Rate in Chlamydomonas Reinhardtii. Plant J. 2011, 66 (5), 770.

Supplementary Table 2: List of Transcription factors used in this study; Family info was adapted from Pérez-Rodríguez et al. (2010) and Jin et al. (2017)

Supplementary Table 3: Edge list representation of the consensus network with mean ranks resulting from Borda count election method (Borda (1781), Marbach et al. (2011)) as edge attributes.

Supplementary Table 4: Top 10 predicted regulators of qE genes in the consensus GRN. Regulation strength was calculatedd based on the edge rank in the consensus network.

Supplementary Table 5: Edge list representation of the PHOT GRN with importance score from GENIE 3 (Huyn-Thu et al., 2010) as edge attributes.

Supplementary Table 6: Top 10 predicted regulators of qE genes in the PHOT GRN. Regulation strength was calculated based on the edge weight assigned by GENIE3.

Supplementary Table 7: List of genes putatively involved in photoprotection used for regulator prediction

Supplementary Table 8: List of genes putatively involved in CCM used for regulator prediction.

Supplementary Table 9: PCR primers for CLiP mutant validation and complementation in this study

Supplementary Table 10: RT-qPCR primers for all genes in this study